DOCUMENT RESUME

ED 204 328 SP 018 607

AUTHOR Hord, Shirley M.; And Others
TITLE Implications of Experience in Teacher Education Program Follow-Up Studies for Future Work. Research and Development Agenda in Teacher Education Colleague Report.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
PUB DATE Apr 79
NOTE 126p.; Proceedings of a symposium at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 11, 1979).

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Beginning Teachers; Educational Research; Feedback; *Followup Studies; *Formative Evaluation; Higher Education; Longitudinal Studies; *Program Evaluation; *Research Methodology; Teacher Attitudes; *Teacher Education Programs; Teacher Employment; Teacher Orientation

ABSTRACT

These papers describe recent techniques used in followup studies of teacher education programs and suggest issues which may evolve from or be enhanced by the study findings. They provide a source of information about current efforts in planning the evaluation of teacher education programs. The first paper describes a study which was designed to document the experiences of first year teachers. The paper draws on the findings to suggest important considerations for future studies of teacher education programs with the same goal and focus. It is suggested in the second paper that teacher education followup studies, although usually carried out to maintain or improve program quality, can have more expanded utility and can address other critical issues. The third paper examines reasons why formative evaluation is needed and discusses issues that must be faced if formative evaluation efforts are to be effective and programs regenerative. In the fourth paper, it is related how, in 1973, Tennessee Technological University developed and implemented a longitudinal model for gathering systematic data from the graduates of its teacher education program. The fifth paper examines critical findings from the Western Kentucky University teacher preparation evaluation program on teacher entry, retention, and change. Included in this collection are comments from two discussants at the symposium. (JD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 261328

# RESEARCH AND DEVELOPMENT AGENDA IN TEACHER EDUCATION COLLEAGUE REPORT

## IMPLICATIONS OF EXPERIENCE IN TEACHER EDUCATION PROGRAM FOLLOW-UP STUDIES FOR FUTURE WORK

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Walton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Research & Development Center for Teacher Education

The University of Texas at Austin

UTR&D

SP 018 607

# IMPLICATIONS OF EXPERIENCE IN TEACHER EDUCATION PROGRAM FOLLOW-UP STUDIES FOR FUTURE WORK

Symposium
Presented at the annual meeting of the
American Educational Research Association
San Francisco, April 11, 1979

TABLE OF CONTENTS

INTRODUCTION . . . . . 1

Shirley M. Hord, Research and Development Center for Teacher Education, The University of Texas at Austin

SOME FEEDBACK IS BETTER THAN OTHERS: IMPLICATIONS OF A STUDY OF FIRST YEAR TEACHERS FOR THE FOLLOW UP OF TEACHER EDUCATION GRADUATES . . . . . 5

Kevin Ryan, The Ohio State University

FOLLOW-UP STUDIES AND RESEARCH ON TEACHER SELECTION . . . . . 25

H. D. Schalock, Teaching Research Division, Oregon State System of Higher Education

FORMATIVE EVALUATION EFFORTS IN A TEACHER EDUCATION PROGRAM . . . . . 29

James M. Cooper and Howard L. Jones, University of Houston

FOLLOWUP STUDIES: PROVIDING THE THRUST FOR PROGRAM IMPROVEMENT . . . 59

Jerry B. Ayers, College of Education, Tennessee Technological University

CRITICAL FINDINGS FROM THE TEACHER PREPARATION EVALUATION PROGRAM: TEACHER ENTRY, RETENTION, AND CHANGE . . . . . 70

Ronald D. Adams, Western Kentucky University

COMMENTS FROM THE SYMPOSIUM DISCUSSANTS . . . . . 109

Karen B. Kepler, Teachers College, Columbia University

Robert F. Peck, University of Texas at Austin

# INTRODUCTION

Shirley M. Hord
Research and Development Center for Teacher Education
The University of Texas at Austin

There is at present evidence of accelerated interest in teacher education program follow-up studies. This may be due to the impact of mandates from NCATE to meet accreditation standards or perhaps to internal institutional pressures to show evidence of program effectiveness. For whatever reason, it appears that evaluation of teacher education programs will be done. Fortunately, there are resources available. The authors of the papers included in this set are currently receiving several requests each week for information on how to/why do follow-up studies. Since these individuals have been involved in such studies for the past five-six years, they have gained considerable experience and expertise.

These six authors are representatives of a consortium which was formed in April, 1978. At that time, teacher educators, in both large state-supported universities and small regional colleges, who had been and who continue to be heavily involved in designing and implementing follow-up studies met to share experiences. This initial meeting at The University of Texas at Austin was supported and hosted by the Research and Development Center for Teacher Education. As representatives from this consortium, two associate deans, a department chairperson, a director of educational research, a director of teacher education programs in a state system for higher education, and a professor in a teacher education program report on their research in follow-up studies of their teacher education programs. Each paper is structured to report in two ways:

Sharing Studies of Teacher Education Programs. The presenters briefly share the important "how-to-do-its" of designing and managing quality studies: information about research questions, study design, selection of measures, concerns about budget and staffing, approaches to data collection and data processing, managing the study, and utilizing feedback.

Using Results as the Cornerstone of Renewal. In addition to describing their work, each author identifies an important implication which now focuses attention for future work. How this new issue may be addressed through findings from their follow-up studies is described.

In the paper, SOME FEEDBACK IS BETTER THAN OTHERS: IMPLICATIONS OF A STUDY OF FIRST YEAR TEACHERS FOR THE FOLLOW UP OF TEACHER EDUCATION GRADUATES, Kevin Ryan, The Ohio State University, describes a study which was designed to document the experiences, the "life space," of first year teachers. The paper draws on the findings to suggest a number of important considerations pertinent to the execution of studies with the same goal and focus.

The eighteen teachers studied ranged in subject area, the teaching setting, grade level which they taught, and on other dimensions. Further, they were graduates of nine training institutions. They were interviewed on five occasions and observed regularly throughout the duration of their induction year.

In designing and carrying out follow-up studies to assess how well a program has prepared its graduates, a number of issues uncovered by this study should be considered. First, the teaching performance of the first year novice was markedly unstable, erratic and strongly influenced by their private lives. This finding has obvious implications for the timing, frequency, and duration of follow-up activities. Second, teachers' subjective judgment about their own skills and performance often differed widely from that of observers; an evaluation based on self-report may therefore be of questionable value. Similarly, the criterion of continued employment (for a second year) seemed to bear little relationship to actual teaching skills, casting doubt on the value of this criterion for the evaluation of a successful teacher preparation program. Finally, the context of each subject's situation appeared to be unique, different enough to preclude assessment of its effects on teaching. The study, in fact, suggested that ultimate teaching success is strongly influenced by initial placement. The author urges caution in attempting to determine the quality of a teacher training program without consideration of the contextual factors operant during the induction phase.

H. D. Schalock, Teaching Research Division, Oregon State System of Higher Education, suggests in FOLLOW-UP STUDIES AND RESEARCH ON TEACHER SELECTION that teacher education follow-up studies, although usually carried out to maintain or improve program quality, can have more expanded utility and can address other critical issues. The attainment of this goal is addressed in this paper by a focus on teacher selection follow-up research.

Research on teacher selection should: (1) be essentially predictive in nature, with the concern being to determine who is likely to have most success in achieving desired learning outcomes with particular groups of children under particular learning conditions; (2) be longitudinal in design, with prospective teachers studied throughout their initial preparation program, and through the first three to seven years of teaching; (3) take into account the impact of the setting in which teaching occurs; (4) obtain measures of effectiveness at various stages of professional development; and (5) control variables likely to interact with predictive measures.

If the above conditions were met by incorporation into ongoing teacher preparation programs and the establishment of continuing programs of graduate follow-up, the benefits are obvious. A low-cost, functional laboratory would not only exist for research on teacher selection, but, by the manipulation of program content/operation, a context could also be established for research on teacher preparation. Further, larger scale studies could be cheaply replicated, yielding an empirical knowledge base about both teacher selection and preparation.

James M. Cooper and Howard L. Jones, University of Houston, report on FORMATIVE EVALUATION EFFORTS IN A TEACHER EDUCATION PROGRAM. Starting with pilot efforts in 1971, the University of Houston has been attempting to develop a formative evaluation system for a competency-based teacher education program. In the subsequent years, management team members within the college increased their efforts to generate a workable system for the identification of the satisfaction and successes of its present and former teacher education students and the degree to which the program has implemented twelve instructional characteristics believed to be essential to the operation of an effective program. Data resulting from these inquiries and procedures followed are reported in this paper.

The most salient point made by the authors is the pressing need for formative use of evaluation results. Past efforts of this type have tended to be minimal, sporadic and generally informal. This paper examines reasons why formative evaluation is needed and discusses issues that must be faced if formative evaluation efforts are to be effective and programs regenerative. Among these issues are: managerial designs and operations that facilitate data collection and usage in large and small teacher education programs; strategies for enhancing the probability that relevant data will be collected; data collection and analysis procedures; types of data that are likely to be relevant and persuasive; and problems of inertia and reluctance to change.

FOLLOW-UP STUDIES: PROVIDING THE THRUST FOR PROGRAM IMPROVEMENT, presented by Jerry B. Ayers, Tennessee Technological University, relates how in 1973 Tennessee Technological University developed and implemented a longitudinal model for gathering systematic data from the graduates of its teacher education programs. The model has served as a base for meeting the evaluation standards of such organizations as the National Council for the Accreditation of Teacher Education. The model has application at other institutions desiring to implement programs for the improvement of their teacher preparation programs.

The application of this model at Tennessee Tech entered its sixth year in 1978-79 with the specific objectives of evaluating the graduates in order to provide information for program improvement at the university. During the period, six distinct groups of subjects have participated in the study, i.e., graduates from the bachelor's and master's level programs for the period 1973 through 1978. Detailed information has been collected on each subject from four distinctly different sources: 1) self-evaluative data; 2) principals and supervisory evaluations; 3) satisfaction of students of the graduates; and 4) data from independent trained observers.

The major findings of the study have indicated that each group of first year graduates have changed little over the five year period for which detailed analyses have been completed. Also, few changes have been noted in the subjects who have been in the study for as much as five years. The graduates have exhibited many of the characteristics of good teachers; however, when compared with national samples, they are functioning near the mean. Even though there have been few apparent changes in the graduates during the period of the study, much valuable information has been gained that provides for ongoing improvement of teacher education programs at the university, an issue of critical interest. How the findings address the important consideration of program modification is highlighted in this paper.

Ronald D. Adams, Western Kentucky University, examines CRITICAL FINDINGS FROM THE TEACHER PREPARATION EVALUATION PROGRAM: TEACHER ENTRY, RETENTION, AND CHANGE. The Teacher Preparation Evaluation Program (TPEP) was implemented at Western Kentucky University in the spring of 1972, to collect objective, quantifiable data on graduates from teacher education programs. More than 250 variables were collected from multiple sources on each participant at four stages: during student teaching and at the end of the first, third, and fifth years of teaching. Data has been collected on more than 300 participants to date.

In part, the research results indicate that secondary teachers who enter teaching

differ in certain characteristics from those who do not enter teaching; elementary teachers who remain in teaching after three years have been found to differ from those who leave the profession after one year of experience. In addition, the evidence shows that teachers' classroom behavior changes as more experience in teaching is gained.

Emanating from these findings are two critical issues: (1) the entry and retention of teachers in the profession, and (2) change in teacher behavior over time. Utilizing the TPEP findings and resultant data base, future work considerations and decision-making by faculty and administrators who are concerned with teacher education may be shaped by the impact of these important issues.

These five papers were reported in a symposium at the annual meeting of the American Educational Research Association in San Francisco. They reveal recent techniques utilized in follow-up studies of teacher education programs and suggest issues which entertain the reflection of study findings in order to initiate new endeavors. Therefore, the papers provide to those involved in or planning teacher education program evaluation a source of information about current efforts. The findings from such studies are diverse and significantly applicable for many purposes. An array of these issues and implications evolving from study results is explicated.

A significant contribution to the symposium presentation and to this set of papers was made by two discussants, Karen B. Kepler, Teachers College, Columbia University, and Robert F. Peck, University of Texas, Austin. Their insightful comments represent the thoughtful and reflective consideration these two scholars brought to their reading and critiquing task. Their remarks add substantially to this small volume; we are indebted to them for their interest and willingness to participate in a reactor capacity. Their presentations follow the text of the five papers.

Persons interested in additional information on follow-up studies of teacher education programs may obtain a volume, containing the ten papers shared at the April, 1978, consortium, entitled,

Teacher Education Program Evaluation and Follow-Up Studies:
A Collection of Current Efforts

Shirley M. Hord and Gene E. Hall, Editors

by writing to:

Communication Services
Research and Development Center for Teacher Education
Education Annex 3.203
The University of Texas at Austin
Austin, Texas 78712

# SOME FEEDBACK IS BETTER THAN OTHERS: IMPLICATIONS OF A STUDY OF FIRST YEAR TEACHERS FOR THE FOLLOW UP OF TEACHER EDUCATION GRADUATES

by
Kevin Ryan
Professor of Education
The Ohio State University

Paper presented at the 1979 American Educational Research annual meeting, San Francisco, California, April 11, 1979.

DRAFT

Since the early 1930's, when E. L. Thorndike (1932, 1935) and his colleagues reported the results of their experiments, educators have shown a strong interest in the principle of feedback in learning. In one experiment, Thorndike showed how feedback on performance aided blindfolded subjects draw a line of a specified length. In another experiment, with Lorge, subjects learned how to toss a ball at a target which they weren't able to see. Again, feedback was crucial to learning.

Preservice teachers are taught the value of feedback for learning in educational psychology courses and the concept is built into their programs, in such activities as microteaching and student teaching. However, for a variety of reasons, not the least of which is limited resources, teacher education programs have used feedback for their own learning and improvement only on a limited basis. Having tossed their graduates out of the schools, programs have rarely followed up their graduates

to see if their training is on target. As a result, teacher educators use only the most impressionistic knowledge of results with which to evaluate and revise their programs.

In recent years the Nation Council for the Accreditation of Teacher Education (NCATE) and other agencies have been urging teacher education institutions to conduct follow up studies of their graduates. It is the author's impression that what little follow up is carried out is done by means of paper and pencil questionnaires and occasionally an accompanying on-site visit. This paper is addressed to these follow up study practices, particularly as they pertain to first year teachers.

Normally when a new teacher begins her first teaching assignment she is cut off, both physically and psychologically, from the college or university where she trained. She is on her own. As a first year teacher, she may receive two or three classroom visitations by and conferences with administrators. However, she rarely receives any special help to aid in the transition from student of

teaching to teacher. Based on the belief that the first year of teaching is a year of intense learning and that it is pivotal to the professional development of the teacher, a group of seven of us* designed and executed a study of first year teachers. The purpose was not to follow up on the graduates of one program. Nor was it to test and develop methodologies for following up on graduates of teacher education programs. Our purpose was to find out what happens in the life space of first year teachers. In the words of Frances Fuller and Oliver Bown (1975), we are attempting to find out, "what's out there?" We wanted to capture as much of the experience of first year teachers as we could and chart the dimensions of that experience. On the other hand, the purpose of this paper is to draw out from the experience of our study directions and suggestions which may be of use to those working on the problem of how to follow up graduates of education programs, particularly during their initial year after graduation

*Dr. Jane Applegate, Mr. Randall Flora, Dr. John Johnston, Dr. Thomas Lasley, Dr. Gerald Mager, Dr. Katherine Newman and Dr. Kevin Ryan.

DESCRIPTION OF FIRST YEAR TEACHER STUDY

Our study attempted to find out what happened in the life space of eighteen first year teachers. We developed five questions which guided our research. They are:

1. What are the first year teacher's perceptions of himself or herself?
2. What are the first year teacher's perceptions of himself or herself as a teacher?
3. What happens in the classroom and in other parts of the school that the new teacher perceives as successes or as failures?
4. What are the new teacher's perceptions of students, administrators, fellow teachers, their teacher educators, and parents?
5. What changes occur in these perceptions that in the lives of first year teachers or the course of the year?

We decided that for the type of study we had in mind, we needed eighteen first year teachers. We wanted these teachers to be teaching throughout the grades from kindergarten to twelfth grade in a variety of settings (urban, suburban, rural and public, parochial and private schools) to reflect the distribution of males and females in the teaching population, to be teaching in a variety of areas from academic to vocational, from regular classrooms to special education classes. We also hoped to include a representative number of minority group teachers. We were, in general, successful on all accounts, except two. Although we had urban schools, we had no first year teachers in classrooms serving the extremes of the urban poor. Second, although we asked minority group teachers to participate, none volunteered. In order to get our sample of eighteen volunteer teacher-subjects, we had to ask thirty-six.

The eighteen teachers in our sample came from nine different teacher training institutions and they were teaching in sixteen different

schools. One teacher, a middle-aged trade and industrial teacher taught carpentry for a month and decided to go back to his trade full time. A second teacher left her position after approximately three months, went home and took up another teaching job close to home later in the Spring.

The seven researchers assigned themselves to individual teachers for investigation, with each researcher assigned to certain teachers. These assignments were kept through the academic year, and as the study continued, so have the assignments.

Our method of investigation had two main features: one, five structured interviews scheduled before, during and after the first year. The interviews took anywhere from forty minutes to an hour and a half to complete. These were taped, and then later summarized on cards. Second, we regularly observed the first year teachers, starting with their second day with their classes and extending throughout the year. These two major approaches gave us first, the

first year teacher's view of what she was doing and what was happening to her; and the second, the observations, gave us an opportunity to see her behaviors and to make judgments accordingly.

There was variety in the way the seven researchers interacted with their first year teacher-subject. Individual researchers interacted with different degrees of frequency with the teachers to whom they were assigned. Some kept regular contact over the phone. There was no standardized method of observing classroom behavior.

The researchers came together on a weekly basis to share their perceptions and discuss various aspects of this study.

## ISSUES FOR CONSIDERATION

Even given the limitations on our study, I believe it does suggest issues to those engaged in designing and conducting follow up studies for the purpose of discovering how well a particular program has prepared its teacher graduates.

First, the teaching performance of first year teachers is not very stable. In our study several of the teachers varied their teaching approaches, their attitudes, and their state of emotional balance a good deal over the year. What this suggests is that in a year marked by instability and change, there is no one best time to send out a questionnaire or a classroom observer.

Second, teachers' subjective judgment of their overall success in teaching or their evaluation of their ability at a particular aspect or skill of teaching often varied a great deal from the judgments of the researcher who had been observing them. For example, in our study a few teachers were, in the judgment of the observers skillful and effective right from the beginning. Sometimes, these teachers were very harsh in their judgments of themselves. On the other hand, some of our first year teachers who were pleased with their own performance were, in the judgment of the observers, either in grave trouble or rapidly approaching it. What this suggests is that the first year teacher's

subjective evaluation of her performance may have very little correspondence to her actual performance or level of skill.

Third, the individual teacher's situation, the context of her teaching, is unique and it is difficult to determine its effect on her professional performance. While several of the teachers in our study left us with the impression that they could succeed in many different teaching situations, others seemed to be gravely affected by their special contextual factors. This suggests that the success or failure of a beginning teacher may depend upon her initial placement. If placement is key, then we must be quite circumspect before we attribute much of the teacher's performance to the uniqueness of a particular preservice training program.

Fourth, the morale, energy level and teaching behavior of many of our teachers seemed strongly affected by what was happening in their private lives. Some of the teachers in our study had

marital stresses and strains, another the death of a parent, at least one a disappointing romance, one went through a divorce, another was the victim of a robbery and assault. It seemed that these events were having a major effect on their teaching. Again, the situational variables from outside their classrooms seemed to interfere with the effect of their training program. Further, this suggests that their response to a questionnaire or their observed teaching performance would be affected by these special events.

Fifth, whether or not first year teacher is invited to teach second year is occasionally mentioned as a criterion of judgment of program effect. Our study could cast doubt on the value of this as a criterion. One of our eighteen teachers quit after five weeks. All of the remaining seventeen were offered contracts the second year. In the view of the observers, some of them were quite ineffective and showed little promise as teachers. There seemed to be little relationship between their performance and their employment.

SOME GENERAL QUESTIONS

Our study of first year teachers raises some other, more general, questions for the designer of follow up studies. First, that in order to provide useful information to program developers, the content of a follow up questionnaire should connect with the reality that is preoccupying the new teachers. Our first year teachers appeared to us to be at many different stages of professional development, from the struggle to survive one day at a time to concern for refining rather sophisticated techniques. This suggests that the follow up study have breadth, spanning concerns from the skills of classroom management to the subtleties of concept formation. The evaluation should be sensitive to such issues as the new teacher's struggle for the courage to make evaluative judgments of her students' work, rather than any particular set of skills of evaluation.

Second, and related to the above, the follow up instrument may reflect the assumption that the first year teacher's energies are being

devoted to acquiring and testing new skills and competencies. In fact, as I have suggested elsewhere (1979), the real task of of first year teachers may be simply to become a second year teacher. It may be that the major energies of the first year teacher are devoted simply to learn to fit in with the rest of the staff and to acquire the operational norms of that particular work setting. For many teachers, the motto of the first year may not be "becoming a teacher" but "going along with the system."

Third, people doing follow up studies may need to establish multiple definitions of first year teachers. Our encounter with the first year teachers in and out of our study suggests that multiple induction patterns may be developing. Besides the standard pattern, whereby a college graduate is immediately hired into a first year teaching position, there are other patterns emerging. For instance, being a district-wide substitute for a year, and then being hired as a regular teacher. The individual who has spent a year substituting in a district may be much more knowledgeable

and at home in their classroom and may be much more skilled at what is for many first year teachers their bete noire, classroom control.

Fourth, there may be special sampling problems in follow up studies of first year teachers. If an institution has a sixty percent return rate of its follow up study instruments from its graduates of the previous year who have obtained teaching positions, they might be quite satisfied with that level of return. However, in the forty percent that have not responded, may be a solid portion of those who are so taxed and overextended by their experience, that they simply cannot give attention, time or energy to something that is so far from their personal survival as a follow up questionnaire. The lack of these respondents in the response pool may badly distort the evaluation of program that emerges.

## SOME INCIDENTAL LEARNINGS

In the process of our study we unexpectedly learned some things, which while they were of

interest to us, may or may not be of interest to those doing follow up studies. The first has to do with the reactions of the teacher-subjects to being involved in the study and to the researchers. In the last five years there has been substantial comment about teachers being tired of being used by researchers. The scenario goes something like this. A fast talking university faculty member mysteriously gains access to some classrooms, collects his data, (in the process upsetting teachers and students), and then going off and, either is never heard from again or writes widely distributed articles where the teachers and administrators look like buffoons and/or facists. In our study we found some initial difficulty in finding a sample of eighteen first year teachers, and we, also, wanted to know their reactions to being a part of the study. We asked them twice during the year to respond to a series of questions about their involvement in the study, their relationship with the researcher and, in general, their level of affect toward the study. They were

told that their responses would not be looked at for over a year. When we finally did look at their responses, we were amazed by three things. First, all sixteen teachers responded to both questionnaires requesting their views on this matter. Second, all sixteen had positive feelings about being in the study and their relationship with the researcher. Third, we had trained ourselves as unobtrusive as possible, simply trying to capture what was there, rather than be any type of intervention. Most of them, nevertheless, found us to be a positive part of their first year. I surmise that while we can take some credit for this reaction it is more an indication of the first year teachers professional isolation and loneliness.

Second, a major impact of this study was on the researchers themselves. The study was funded by the individual researchers. It was not part of anyone's formal university load. It was conducted on top of regular assignments.

No one used it as his dissertation study. Nevertheless, it has been seen by the

researchers as a major part of their doctoral programs and their preparation for careers as teacher trainers and educational researchers. In this sense, it has been an ideal training experience.

Third, our data forced us to question Lagana's finding about the "curve of disenchantment" in which first year teachers' attitudes towards students and their work took a steep dive late in September and did not level off until January. Based on self reports, the teachers in our study show strikingly different profiles of satisfaction from the curve of disenchattment and one first year teacher from the other (see figure 1).

Thus far in your first year of teaching you have expressed things that have pleased you or made you feel good about yourself and teaching and things that have upset you or disappointed you. On the time line below would you mark what you consider to be the high and low points in the year for you thus far.

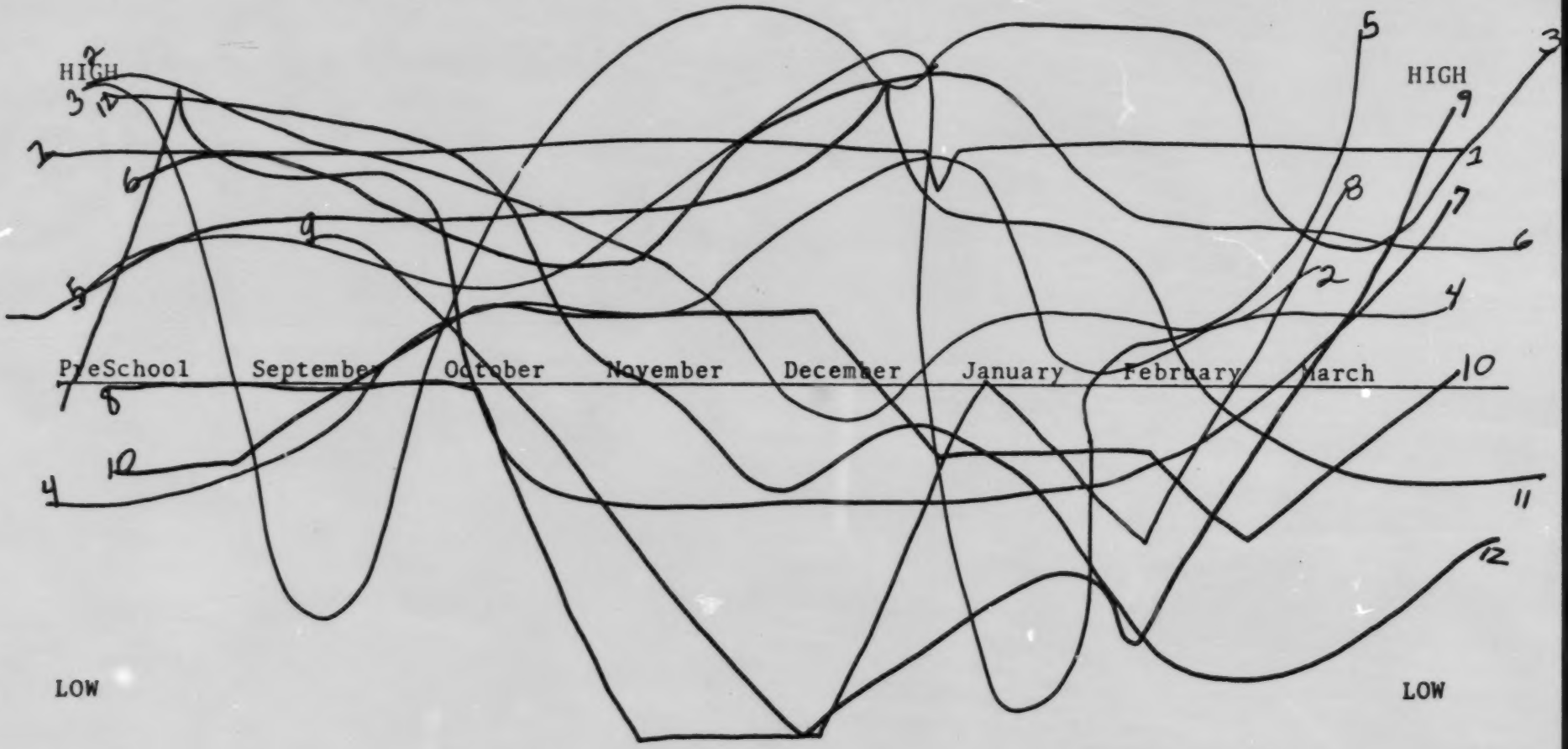



FIGURE 1

## CONCLUSION

Last Spring the National Institute of Education circulated a request for proposals (1978) which contained the staff's views on the induction phase. After reviewing the research, the P stated:

> One result of this effort is a conclusion concurred in by thoughtful observers with a variety of backgrounds and concerns: the conditions under which a person carries out the first year of teaching have a strong influence on the level of effectiveness which that teacher is able to achieve and to sustain over the years; on the attitudes which govern teacher behavior over even a four year career; and, indeed, over the decision whether or not to continue in the teaching profession.

We need to know much more about the first year of teaching and its effects. One important way to do this would be to significantly improve our methods of following up on the graduates of our programs.

## REFERENCES

Fuller, F., and Bown, O. (1975) "Becoming a Teacher" Kevin Ryan (ed.) Teacher Education: The 74th Yearbook of the National Society for the Study of Education (part II) Chicago: University of Chicago Press.

Lorg, I., and Thorndike, E.L. (1935) The Influence of Delay in the Aftereffect of a Connection. J. Exp. Psychol. 18 186-194.

Ligana, J. (1970) What Happens to the Attitudes of Beginning Teachers. Danville, Illinois: Interstate Printers and Publishers, Inc.

National Institute of Education (1978) Request for Proposals (RP-NIE-R-78-0014) Washington, D.C.: Department of Health, Education, and Welfare.

Ryan, K. (1979) Inside the Black Boxes: The Process of Teacher Education. A paper presented at the Research and Development Center on Teacher Education sponsored conference, on Exploring Issues in Teacher Education: Questions for Further Research, January 10-12, 1979. Austin, Texas.

Thorndike, E.L. (1932) The Fundamentals of Learning. New York: Teachers College.

# FOLLOW-UP STUDIES AND RESEARCH ON TEACHER SELECTION[1]

H. D. Schalock
Teaching Research Division
Oregon State System of Higher Education

Follow-up studies in the context of teacher education usually are thought of in terms of quality control (Arnold, et. al., 1977). They are carried out as a means of assuring program quality or guiding program improvement. This way of thinking about follow-up studies has been fostered largely by accrediting agencies, and it tends to dominate the design and use made of follow-up efforts.

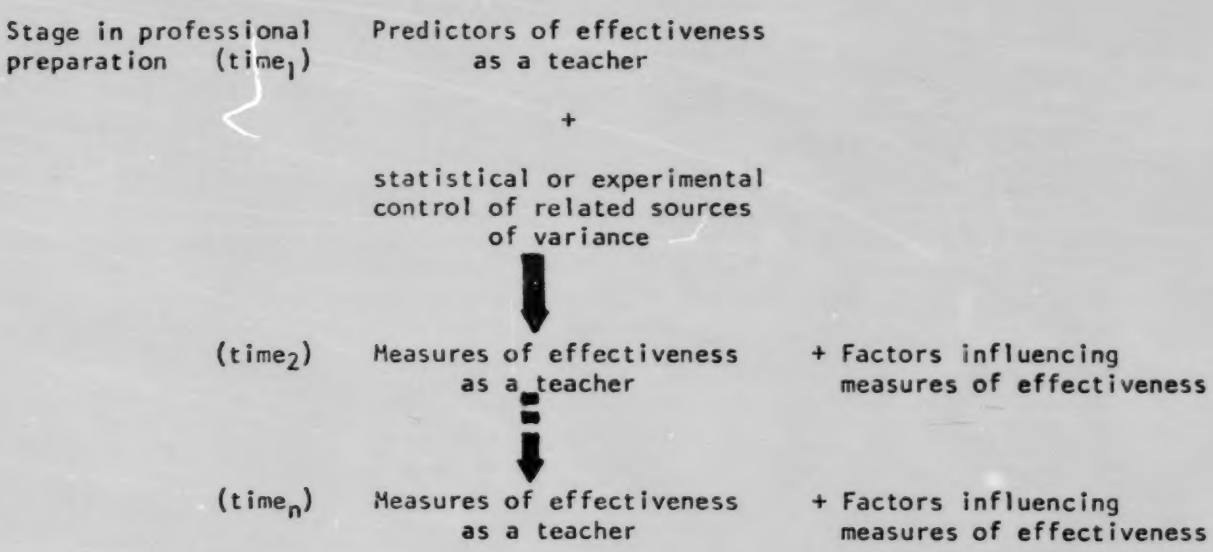
In the few minutes I have today I want to argue that follow-up studies of graduates of teacher preparation programs, if well done, can become the backbone of research on a number of issues that are critical to teacher education. I will limit my remarks this afternoon to the use of follow-up data in research on teacher selection, but follow-up data are equally appropriate for research on the effectiveness of teacher preparation programs or on the cost and benefits associated with alternative approaches to the preparation of teachers. The argument could be made, in fact, that research in any of these areas depends ultimately on follow-up data of some form being collected.

I have argued elsewhere (Schalock, 1979) that research in teacher selection is essentially predictive research, where the predictive concern is one of who is likely to have the most success in achieving desired learning outcomes with particular groups of children under particular learning conditions. Given this view research on teacher selection needs to be longitudinal in design, with prospective teachers being studied throughout their initial preparation programs and through at least their

---

[1] Presented at the meetings of the American Educational Research Association, San Francisco, April 1979.

first three and preferably their first five or seven years of teaching. This view of teacher selection research can be shown schematically as follows:


Stage in professional preparation ($time_1$)
Predictors of effectiveness as a teacher
+
statistical or experimental control of related sources of variance
($time_2$)
Measures of effectiveness as a teacher
+ Factors influencing measures of effectiveness
($time_n$)
Measures of effectiveness as a teacher
+ Factors influencing measures of effectiveness

Central to the framework is the concept of stage in professional preparation, and the related assumption that measures are obtained routinely at each stage to inform decisions about continuation in the profession. If such measures were obtained, and they were taken with the view to being used in research, they would serve the interests of research on teacher selection perfectly. Measures obtained in the early stages of a professional preparation program could be used as predictors of success. Measures obtained in follow-up studies could be used either as dependent measures or predictors of subsequent success. The points in professional development where assessment routinely occurs include: (a) entrance to a preparation program; (b) recommendation to engage in student teaching; (c) recommendation for an initial level of certification; (d) recommendation for job continuation after the first year of teaching; and (e) recommendation for continuing or an advanced level of certification, or tenure.

The framework carries several additional assumptions.

- If measures routinely collected in professional preparation programs were viewed as data to be used in research the quality of the measures taken probably could be improved with little added expense or effort.

- Variables likely to interact with the predictive measures being investigated need to be controlled through statistical or sampling procedures. For example, if knowledge of subject matter or mastery of teaching skills are being studied as predictors of teaching effectiveness, amount of experience in teaching needs to be controlled.

- After entry to a preparation program has been granted assessment at each stage in professional preparation should include a measure of teaching effectiveness. Follow-up studies of graduates of teacher preparation programs should have as their central focus measures of teaching effectiveness.

- Features of the setting in which teaching occurs that are likely to influence the ability of a teacher to effect desired learning outcomes, for example, the characteristics of children taught or the amount of time allocated to a particular subject area, should always accompany measures of teaching effectiveness.

If these four conditions could be met, and more importantly incorporated as part of an ongoing teacher preparation program, research on teacher selection could proceed with relatively little cost beyond that required to operate the preparation program. It would have the built-in advantage of the conditions of high external validity and essentially endless replications.

All of this presupposes, however, an effective and continuing program of follow-up of graduates. With certification agency standards tightening in this regard, and legislative demands increasing for demonstrated quality in teacher preparation programs, this is not an unreasonable supposition.

In many respects what is being proposed is analogous to research laboratories that have been established in other disciplines. If the conditions that have been called for could be implemented it would mean essentially that all of the requirements necessary for a teacher preparation

program to become a functional "laboratory" for research on teacher selection would exist. Moreover, if the content or operation of the preparation program itself could be varied systematically a laboratory would exist for research on teacher preparation as well. Of considerable importance is the fact that such a laboratory could be operated at relatively low cost. Since data would be collected as part of the ongoing preparation program, or as part of the follow-up program on graduates, the most expensive feature of research is eliminated. This, coupled with the opportunity to replicate studies with each new group of prospective teachers entering the program, would permit a line of inquiry to be pursued over a sufficiently long period of time and with a sufficiently large number of replications that trustworthy knowledge actually would have a chance of being produced. This is a condition that has never before existed in teacher education, and if established represents for the first time the possibility that both teacher preparation and selection could move to an empirically based mode of operation.

## References

1. Arnold, D.S., Denemark, G., Nelly, E.R., Robinson, A., & Sagan, E.L. Quality control in teacher education: Some policy issues. Washington, D.C.: American Association of Colleges for Teacher Education, 1977.

2. Schalock, H.D. Research on teacher selection. In D.C. Berliner (Ed.) Review of research in education. Volume 6, 1979, forthcoming.

# FORMATIVE EVALUATION EFFORTS IN A TEACHER EDUCATION PROGRAM

James M. Cooper and Howard L. Jones

University of Houston

A paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1979

Need for Formative Evaluation

Teacher education program evaluation stands at an interesting and critical point today. It is generally agreed that past evaluation efforts have been minimal with respect to both quantity and quality. For many institutions program evaluation means sending out a questionnaire to a sample of program graduates shortly before the next NCATE visitation. While this practice perhaps met the letter of the law with respect to the old NCATE standards, the new 1979 NCATE Standard 6 requires much more in the way of program evaluation. Specifically, the standard states:

> Maintenance of acceptable teacher education programs demands a continuous process of evaluation of the graduates of existing programs, modification of existing programs, and long-range planning. The faculty and administrators in teacher education evaluate the results of their programs, not only through the assessment of graduates but also by seeking reactions from persons involved with the certification, employment, and supervision of its graduates. The findings of such evaluation are used in program modification.

Program evaluation is meant to be an ongoing process, one that aids decision makers in refining their programs. Its greatest power lies in its formative uses.

TenBrink defines evaluation as "...the process of obtaining information and using it to form judgments which in turn are to be used in decision making."[1] The three key elements are information, judgments, and decisions, with the kinds of judgments and decisions needing to be made dictating the information needs. The developers of each program must decide what questions they want to answer before collecting any information. In this

[1] Terry D. TenBrink, Evaluation: A Practical Guide for Teachers, New York: McGraw-Hill Book Co., 1974, p. 8.

way, they ensure that the data they do collect is relevant to their needs; decisions can thus be more validly data-based.

Potential Roadblocks/Problems to Effective Formative Evaluation

Faculty and staff of most institutions of higher education would generally agree that their institutions need more formalized means for collecting formative data about their programs. However, most institutions have loosely organized efforts to collect minimal data. Course and faculty evaluations by students are generally completed after each semester or quarter. In some programs there is a concerted effort to gather summative assessment data at the completion of internship or student teaching.

These same faculty and staff generally would agree that the student and faculty assessments of "during-program" stages is unnecessarily limiting. In a recent survey, it was noted that 80+% of teacher education stitutions valued the gathering of data about their teacher graduates. In the same survey, however, it was noted that 49+% of the institutions were not gathering any follow-up data.[2]

The position presented in this paper is that effective formative evaluation must be made at all levels of the program - before the students start, during the program, and after the graduates start their professional careers. Part of this position is the acceptance of the fact that there are roadblocks and problems that must be solved by any institution before effective procedures can be developed and implemented. Among these problems are the following:

[2] Personal communication from Gary Peer, Assistant Dean of Education, University of Tulsa, to Edell Hearn, Dean, College of Education, Tennessee Technological University, February 28, 1978.

1. Overcoming Inertia and Reluctance

Most faculty, staff and students of teacher education programs have not been engaged in the gathering of formative data about their efforts. How can students be convinced that their suggestions, complaints and even satisfactions will be used by program developers in a useful manner? How can students be convinced that their ideas will be taken seriously? How can faculty members be convinced that formative data is more than "fuel for paranoia?" How can faculty members be helped to deal with the opportunity to improve their offerings rather than looking for another course that is "worse", thereby evidencing the "I'm OK compared to him" syndrome? How can graduates be convinced that it is important to take the extra effort to provide feedback to their alma mater for the purpose of changing elements?

2. What Kinds of Data Should be Gathered?

What data are most useful for the program? Which data are the most timely, persuasive and relevant? General "satisfaction" data are usually diffuse and difficult to use to change programs. Rating scales or similar satisfaction scales really do not permit a "fine tuning" of program offerings, yet these data are the easiest to gather. The gathering of personality characteristics data is usually a stressful effort, yet these data might provide some indication of program effects on students. Are these data worth the psychological and other costs necessary to gather them? Is it important to gather low inference data about the classroom mannerisms of student teachers or graduates? Finally, how much data is enough? Program developers and evaluators must

recognize that over-gathering data can have a deleterious effect on students' satisfactions.

3. Who Will Make the Key Decisions?

Teacher education programs at most institutions are loosely organized, diffuse structures. Accountability and responsibility for individual progress, not to mention the progress of entire programs, are often very nebulous. A key element in formative evaluation is the identification of who is in charge of the evaluation. Who will make the decisions on the selection/development of assessments? Who will administer instruments or collect data within teachers' classrooms? Who will help program builders and implementers interpret the data that are gathered? In short, who owns what decisions and who will have both the responsibility and the accountability of the evaluation effort?

4. What Strategies Will Be Used to Gather the Data?

The successful gathering of useful data takes planning; resources (fiscal and personal) are usually scarce or unavailable. Where these resources come from and the assurance that there is maximum synergy in the collection of data will depend on the viability of planned strategies. Where will the finances come from? Will faculty be provided released time to gather data, especially data from the field? How can assurances be made that no one set of students or faculty is overburdened by data gathering efforts?

These are but a few of the problems and issues that must be considered if formative evaluation of teacher education programs is to work successfully.

Formative Evaluation Efforts at the University of Houston

In recognizing the importance of continuous formative evaluation in teacher education, a task force of faculty members was formed in 1977 to develop and pilot test an evaluation system for the undergraduate offerings of the College of Education at the University of Houston. 1977 was an especially critical year for the College. In 1971 the first experimental CBTE program was launched, serving about 10% of the undergraduate students in the College. In 1974, CBTE was ratified as the one instructional system for the entire college. Thus, three years later (1977) it was considered time to generate a workable system for the collection and use of formative data.

The developed plan for evaluation was an expansion of past program assessment procedures at the University of Houston. Throughout its history of teacher education there have been attempts to determine the satisfaction of teacher education students with their program as well as the success of graduates as teachers. However, recent changes in programs made it imperative to consolidate the informal evaluation processes. Between 1974-77, faculty and staff of the College of Education agreed to a set of programmatic assumptions to guide the development and implementation of their teacher education program. In addition, a list of program characteristics, as well as a set of generic competencies relating to special subject matter knowledge and teaching skills, were identified.

While the assumptions, program characteristics and competencies noted above were agreed to in principle, the level of implementation of these elements was unknown. The purpose of this research effort was to identify how well these elements had been implemented.

Specifically, the initial program evaluation effort had two major goals; each identified by an evaluation sub-system.

Sub-system 1: A discrepancy evaluation between what presently occurs in the undergraduate teacher education program and those programmatic characteristics that should be evident given the agreed upon assumptions about what constitutes an effective instructional system.

Sub-system 2: The identification of the satisfaction and successes of both present and former teacher education students at the University of Houston.

## SUB-SYSTEM 1 - PROGRAM CHARACTERISTICS

In 1975 the faculty of the College of Education formally adopted twelve assumptions regarding instructional system characteristics believed to be essential to the operation of an effective teacher education program. These twelve characteristics formed the foundation upon which the total instructional system was to be based. This section of the paper describes an evaluation effort to determine the extent to which the twelve instructional characteristics which had been adopted in 1975.

The twelve characteristics believed essential to an effective teacher education program are that the program be:

1. Competency based; that is, the competencies to be demonstrated by the student are made explicit, the criteria to be applied in assessing the student's competencies are made explicit, and the student is held accountable for meeting those criteria. The emphasis is on the demonstration, not the acquisition, of those competencies specified as program expectations.

2. Campus-centered and field-oriented; that is, students are provided with opportunities to experience instruction and to demonstrate competence both on campus and in field sites depending upon the nature of the particular instructional or assessment activity, and a determination as to where that activity can most effectively and efficiently take place.

3. Role-model based; that is, a conceptual model of the teacher's role is used as a basis for identifying and specifying those competencies students are expected to demonstrate.

[3] This evaluation effort is reported in more detail in: Wilford A. Weber and James M. Cooper, "Evaluation of Instructional System Characteristics of the Professional Teacher Preparation Program at the University of Houston." in Teacher Education Program Evaluation and Follow-Up Studies: A Collection of Current Efforts, Shirley M. Hord and Gene E. Hall (eds.). Austin: University of Texas R & D Center for Teacher Education, 1978, pp. 49-66.

4. Criterion-referenced; that is, assessment procedures are designed to determine whether or not a student has demonstrated competency at or above the level of mastery specified; the competency of each student is judged on the basis of predetermined criteria and is not determined through a comparison involving other students.

5. Pluralistic; that is, there is conscious acceptance of the notion that no one philosophy of instruction has been proved best; consequently, a variety of divergent views are posited as untested assumptions which the system must test.

6. Humanized; that is, program design and operation recognize the dignity and worth of each individual so that each perceives that he or she is being treated as one of worth.

7. Personalized; that is, the student is provided with instruction which takes into account his or her uniqueness.

8. Modularized; that is, the delivery of instruction is accomplished through the utilization of instructional modules; instructional modules are sets of learning activities--rationale, objectives, prerequisites, pre-assessment procedures, learning alternatives, post-assessment procedures, and remedial procedures--which are intended to facilitate the learner's acquisition and demonstration of a particular competency or set of competencies.

9. Multi-institutional; that is, the design and operation of the teacher education program is a responsibility shared by colleges, public schools, and the organized teaching profession.

10. Systemized; that is, the systems approach is used in program design and operation.

11. Regenerative; that is, the program is an open system capable of continuous revision on the basis of constructive data supplied by sound formative and summative evaluation procedures.

12. A single system with alternatives within it; that is, the program is a single instructional system which constitutes an integrated, comprehensive whole; however, the system does accommodate a divergence of viewpoints within its confines.

Data Needs and Collection Procedures

A sub-group of the evaluation task force generated sets of indicators which one might expect to find if each of the twelve instructional system characteristics had been operationalized. Those indicators which appeared to be most important and for which it appeared feasible to collect data during the spring 1977 semester were agreed upon by the task force.

The indicators about which data were to be collected were of two types: (1) perception; and (2) documentation. It was determined that perceptual data concerning the twelve characteristics should be collected from students, instructors, supervisors, and supervising teachers associated with the program. Several types of data-gathering procedures were considered to be appropriate for this purpose. However, it was decided that these data, given feasibility considerations, would best be collected through the use of questionnaires--a reasonably good means of collecting perceptual data. This decision led to the construction of four questionnaires, one for each of the four groups to be sampled. Additionally, it was determined that data concerning the characteristics might be gained through analyses of various program documents, policy statements, and instructional materials. It was decided that the use of checklists might be an effective means for this purpose since a checklist can be designed to provide fairly easily obtained, objective information as to whether or not a given characteristic is present.

Questionnaires were mailed to all supervising teachers and student teachers toward the end of the spring semester. At about the same time, questionnaires for the instructors and university supervisors were distributed by instructors in selected classes so that a copy might reach every student enrolled in the PTPP. Questionnaires were returned by way of the same channels.

In order to obtain the program documents which were to be analyzed, the sub-group provided the associate dean for undergraduate studies with a list of each of those documents which was needed. The associate dean, in turn, requested various individuals with Professional Teacher Preparation Program responsibilities to provide the specified documentation. These documents were to be sent to the associate dean, who passed them on to the sub-group for analysis.

Some Limitations

Before briefly describing a few of the study's findings, it is perhaps wise to consider several of the factors which may limit those findings. First, the findings reported here are valid to the extent that the indicators chosen to represent the program characteristics do in fact reflect those characteristics. Second, the accuracy of the findings is influenced by the extent to which the data chosen for analysis represent the selected indicators. Third, the findings and conclusions are limited by the fact that the return rate on the questionnaires and the requested documentation was only moderately high. Fourth, the findings are limited by the reliability and validity of both the questionnaires and the analysis of documents; no attempt was made to prepare statistical estimates of their validity and reliability. And fifth, the findings are limited to the extent that questionnaire responses and program documentation were accurate and honest.

Space precludes reporting more than just a few of the findings from the questionnaire data and the document analyses. Several of the more interesting items are described here to provide examples of the kinds of data obtained and the manner in which they were treated.

Questionnaire Data. The first twelve items on each of the four questionnaires which were administered to the students, instructors, supervisors, and supervising teachers in this study consisted of statements which presented each of the

program's assumptions concerning an effective instructional system. Using a nine-point scale, with one being strongly disagree and nine being strongly agree, respondents were asked to indicate the degree to which they felt that the program had been consistent with each assumption. For example, the first statement on each of the four questionnaires was concerned about the extent to which respondents perceived the program to be competency based.

The statement to which respondents were to react said: "The program is competency based; that is, the competencies students are asked to demonstrate are made explicit, the criteria applied in assessing the competencies of students are made explicit, and students are held accountable for meeting those criteria.

Table 1 presents the results obtained from students, instructors, supervisors, and supervising teachers for those first twelve items. Mean score data suggest that all four groups agreed that the program had operated in a manner which was consistent with the stated assumptions. Mean scores ranged from a low of 5.13 to a high of 7.85. Student mean scores ranged from 5.99 (personalized) to 7.29 (campus-centered and field-oriented) with the other ten scores falling between 6.10 and 7.04, a rather small range.

Instructor mean scores ranged from 5.13 (personalized) to 7.64 (campus/field); overall they were a bit lower than student mean scores with five scores ranging from 5.13 to 5.96, and another six scores ranging from 6.02 to 6.71. Supervisor mean scores ranged from 5.38 (role-model based) to 7.47 (humanized); four of these ranged from 5.38 to 5.78, while seven ranged from 6.23 to 6.96. Overall, the mean scores obtained from supervising teachers were the highest; they ranged from 6.33 to 7.85.

These results were encouraging. While the data suggest that there was much room for improvement, they also indicated progress. The data from university personnel--instructors and supervisor, those most closely involved in the program's development--indicated that they agreed the program was operationalized in a way

Table 1

Summary of Student, Instructor, Supervisor, and Supervising Teacher Questionnaire Data for All Twelve Program Characteristics, Variables 1 through 12. Items 1 through 12

| Program Characteristic | Students | | | Instructors | | | Supervisors | | | Supervising Teachers | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| Competency Based | 510 | 6.55 | 1.99 | 44 | 6.55 | 1.66 | 23 | 5.61 | 1.88 | 213 | 6.33 | 2.31 |
| Campus/Field | 503 | 7.29 | 1.70 | 44 | 7.64 | 1.66 | 23 | 6.65 | 2.08 | 197 | 7.26 | 1.88 |
| Role-Model Based | 493 | 6.24 | 1.96 | 42 | 5.88 | 1.99 | 21 | 5.38 | 2.22 | 197 | 6.80 | 1.94 |
| Criterion-Referenced | 496 | 6.49 | 2.11 | 43 | 6.02 | 2.45 | 23 | 6.96 | 2.16 | 206 | 6.77 | 2.20 |
| Pluralistic | 503 | 7.04 | 2.10 | 43 | 6.05 | 2.25 | 23 | 6.91 | 2.15 | 206 | 7.42 | 1.92 |
| Humanized | 509 | 6.92 | 2.04 | 44 | 6.41 | 2.43 | 23 | 7.47 | 1.90 | 212 | 7.85 | 1.72 |
| Personalized | 507 | 5.99 | 2.37 | 45 | 5.13 | 2.70 | 23 | 5.78 | 2.65 | 196 | 6.95 | 2.21 |
| Modularized | 501 | 6.98 | 1.95 | 45 | 6.71 | 1.97 | 22 | 6.23 | 2.14 | 190 | 6.52 | 2.14 |
| Multi-Institutional | 510 | 6.96 | 1.80 | 45 | 5.96 | 2.54 | 23 | 6.39 | 2.54 | 212 | 7.73 | 1.79 |
| Systemized | 508 | 6.33 | 2.05 | 45 | 6.11 | 2.35 | 23 | 6.65 | 2.04 | 197 | 7.17 | 1.83 |
| Regenerative | 491 | 6.35 | 2.09 | 45 | 5.64 | 2.59 | 22 | 6.77 | 2.07 | 191 | 7.35 | 1.77 |
| Single System | 505 | 6.10 | 2.03 | 45 | 5.24 | 2.52 | 22 | 5.73 | 2.12 | 193 | 6.90 | 1.99 |

which was consistent with the specified assumptions. However, they were the groups which perceived the greatest lack of congruence between program intent and program reality. They seem to have said that "we're getting there but we're not there yet." Student data suggest that students also agreed that the program was reasonably consistent with the assumptions on which it was based. However, the data also suggest areas needing improvement. On the other hand, data from the supervising teachers seem to indicate that they felt the program was well on its way to "getting where it wanted to go." One might speculate that their views are colored by their experiences with other teacher education programs, including those in which they were undergraduates.

In summary, then, the data from the first twelve items on each of the four questionnaires, as presented in Table 1, suggest that while progress had been made in operationalizing the instructional system characteristics there existed much room for improvement.

In an attempt to gain a greater understanding of student perceptions, student questionnaire data were analyzed according to the respondent's certification program (elementary education; secondary education; art education; health and physical education; or music education), and his or her standing (Phase I, II, III, or IV). This type of analysis made it possible to see if sub-groups of students in the program perceived it similarly or differently.

The first twelve items of the questionnaire provided data relevant to broad areas, while the other items focused on specific aspects of the program, particularly the methods courses, foundations of education course, introductory courses, and the multicultural course. For example, the data reported in Table 2 indicate that the foundations of education course (FED 361) was perceived by students to have been considerably less competency based than the other three. Similarly, data reported in Table 3 indicate that the foundations of education

Table 2

Summary Data from Student Questionnaire for Characteristic 1
(Competency Based)
Items 36 through 51

| Item | Course | N | Mean | SD |
|---|---|---|---|---|
| 36 | FED 361 | 391 | 5.27 | 2.87 |
| 40 | FED 361 | 393 | 5.15 | 2.85 |
| 44 | FED 361 | 386 | 5.31 | 2.85 |
| 48 | FED 361 | 385 | 5.77 | 2.80 |
| 37 | C&I 362-333 | 476 | 7.20 | 2.06 |
| 41 | C&I 362-333 | 480 | 6.00 | 2.08 |
| 45 | C&I 362-333 | 480 | 6.86 | 2.20 |
| 49 | C&I 362-333 | 462 | 7.08 | 2.11 |
| 38 | C&I 335 | 210 | 7.04 | 2.05 |
| 42 | C&I 335 | 205 | 6.71 | 2.31 |
| 46 | C&I 335 | 208 | 6.75 | 2.32 |
| 50 | C&I 335 | 201 | 6.90 | 2.21 |
| 39 | Methods | 342 | 6.94 | 2.09 |
| 43 | Methods | 337 | 6.77 | 2.09 |
| 47 | Methods | 330 | 6.84 | 2.01 |
| 51 | Methods | 323 | 6.88 | 2.04 |

course was also perceived to be considerably less modularized than the other three groups being rated. These findings generated great concern regarding the structure of the foundations of education course.

Table 3

Summary Data From Student Questionnaire for Characteristic 8
(Modularized)
Items 114 through 117

| Item | Course | N | Mean | SD |
|---|---|---|---|---|
| 114 | FED 361 | 360 | 3.57 | 2.92 |
| 115 | C&I 362-333 | 454 | 7.09 | 2.20 |
| 116 | C&I 335 | 194 | 7.17 | 2.05 |
| 117 | Methods | 310 | 6.22 | 2.57 |

Document Analyses

In addition to the collection and analysis of perceptual data, the evaluation--as noted earlier--included an analysis of various pertinent program documents. As with the questionnaire data, the intent was to look for evidence that the program characteristics which had been specified as assumptions had been operationalized. The sorts of documents which were requested included:

1. Descriptions of all instructional modules used in PTPP. These were examined for evidence of competency based instruction, modularized instruction, and criterion-referenced assessment, for example.

2. Descriptions of faculty, program area, and department grading policies and copies of all course grades given to students during the fall 1976 semester. These were examined for evidence of criterion-referenced assessment procedures and grading, for example.

3. Program policy statements. These contained descriptions of formally adopted goals, assumptions, and required student competencies. These were examined for evidence of a number of the characteristics.

The major findings from the document analyses was that less than half of the PTPP courses were modularized and of those which were, the modules were often deficient in terms of the components they were to contain. Table 4 presents the data for C&I 362: Introduction to the Profession of Teaching, a course for which the findings were rather typical.

Data Utilization

Findings from the module descriptions analyses were made available to program managers in June 1977. They used the findings to begin to identify priority areas needing improvement. As a result, a task force was formed in the summer of 1977 and charged with responsibility for revising instructional modules for C&I 362 and 333. The products of that task force, thirty-nine modules, were designed so as to ensure that each contained all of the components instructional

Table 4

An Analysis of the Instructional Modules: C&I 362.
Introduction to the Profession of Teaching

| Module | Rationale | Objectives | Prerequisites | Pre-Assessment | Learning Alternatives | Post-Assessment | Remediation |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 2 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 3 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 5 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 6 | 1 | 2 | 2 | 2 | 2 | 2 | 0 |
| 7 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Code: 0 = Component not present.
1 = Component present; did not meet criteria.
2 = Component present; met criteria.

modules were expected to contain, and to ensure that all components met the criteria which had been established.

A summary report of the evaluation findings was made available for distribution to all faculty involved in the PTPP. Additionally, program managers held individual conferences with faculty members in various program areas to share data relevant to those program areas.

## SUB-SYSTEM-2 SUCCESS AND SATISFACTION OF STUDENTS

Program developers face a dilemma in the determination of the success of program graduates. Program graduates, both those completing student teaching and in-service teachers, should have a positive impact on their pupils. By assessing this impact using standardized or other tests, an indicator of teacher success would be evident. This "accountability" model is considerably easier to discuss than it is to implement, however. The variance of school settings, class sizes, available teaching aids, pupil aptitudes and motivations all deter true accountability evaluations.

A second possibility of evaluating PTPP products - student teachers and in-service teachers - is that of collecting data on the teachers as they teach. This step would call for classroom observations and the use of high or low inference instruments to evaluate the teachers' interactions with pupils.

Difficulties with the implementation of this effort are also evident. The time needed for actual classroom data collection is only one difficulty. Furthermore, the UH PTPP program has placed a high positive value on the use of clinical supervisory modes. In such modes, supervisors work with student teachers and School-Based Teacher Educators (cooperating teachers) in assessing, in a formative mode, the processes of the student teachers. By implementing what could be perceived as a "big brother" approach to data collection, it was felt that much of the positive affect associated with the clinical supervision mode would be damaged.

To collect data on the success of students it was decided that a high-inference rating scale would be used. The scale would be administered to student teachers and their supervisors. In addition, the School Based Teacher Educator in whose classroom the student teacher worked would be asked to rate the student teacher on the scale items.

In April, 1977, near the completion of the student teaching experience, all student teachers, their supervisors and School Based Teacher Educators were administered a rating scale to assess the success of the student teacher.

The instrument consisted of a series of 1-5 Likert scales on which the rater identified his/her ratings. The items reflected 15 of the 18 generic[4] teaching competencies of the undergraduate program - competencies that all students in all programs are expected to master. Because a few competencies are quite broad, these competencies were assessed in two or more items. For example, Competency Eight (Promotes Effective Patterns of Communication), was subdivided into four items - Communicates with Students; Communicates with Other Teachers; Communicates with Supervisors; Communicates with Parents).

Furthermore, to identify some level of perceived importance of each of the competencies, raters were asked to identify (again on a 1-5 scale) the importance that they associated with the statement. In cases where raters felt they could not respond to an item they were instructed to indicate "Can't Rate".

In essence, raters (students, supervisors, School Based Teacher Educators) indicated their perception of student teacher's demonstration of program competencies, as well as identifying their personal feelings about the value of each

[4]Three additional competencies have been added to the curriculum of the PTPP program: (16) incorporation of career education concepts into instruction: (17) knowledge and skills pertaining to mainstreaming; (18) demonstration of adequate content knowledge.

competency statement. Anonymous responses were received from 191 student teachers, 22 University supervisors of the student teachers, and 223 School Based Teacher Educators.

PTPP Success Study - Student Teachers

Table 5 reports statistical data for each of the Perceived Importance and Student Teacher Success parts of the PTPP Questionnaire. In viewing the entire set of data several elements are evident:

1. There appears to be general agreement with the importance of program competencies. Not one competency item was rated below 4.124 by any of the three rating groups. Thus, there appears to be a general consensus of programmatic thrusts. This should not be surprising considering the effort spent in acquiring SBTE and faculty input into the program. However, it is rewarding to see the agreement.
2. In all cases of the rating of student performance, student teachers generally rated themselves higher than their corresponding SBTEs and University supervisors. It is difficult to account for this since students were assured anonymity and were told that their reports would not count on grading, etc.
3. In viewing each of the tables, it appears that Elementary Education SBTEs and student teachers were consistently rating the competencies and student performance higher than the University supervisor. Again, it is difficult to account for this.

Because of the difficulties in obtaining truly reliable ratings across groups on student performance and perceived importance of the competency statement, an additional step was taken. The mean ratings for each group on each of the scales were ranked within group. For example, SBTE mean ratings of all

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | tem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 15.0 | 4.488<br>18 | 4.727<br>13(T)* | 4.611<br>14(T)* | 1. Apply knowledge of human growth and development in planning for instruction. | 4.167<br>21 | 3.727<br>20 | 4.033<br>18 | 19.7 |
| 9.7 | 4.699<br>10 | 4.909<br>6(T)* | 4.617<br>13 | 2. In planning for instruction apply knowledge about how students learn. | 4.154<br>22 | 3.682<br>21 | 4.037<br>17 | 20 |
| 20.3 | 4.326<br>20 | 4.409<br>19 | 4.140<br>22 | 3. Write clearly stated lesson and unit objectives. | 4.410<br>12 | 4.046<br>11 | 4.192<br>11 | 11.3 |
| 16.7 | 4.571<br>13 | 4.454<br>18 | 4.426<br>19 | 4. Construct understandable and useful lesson and unit plans. | 4.484<br>8 | 4.182<br>9 | 4.252<br>7 | 8 |
| 19 | 4.535<br>16 | 4.364<br>20 | 4.351<br>21 | 5. Implement instruction that is based on designed unit and lesson plans. | 4.436<br>11 | 4.227<br>7 | 4.242<br>8 | 8.7 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | tem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 17.7 | 4.479<br>19 | 4.476<br>17 | 4.468<br>17 | 6. Construct and use tests and other instruments to monitor student progress. | 4.403<br>13(T)* | 3.882<br>15 | 3.984<br>20 | 19.3 |
| 14.3 | 4.532<br>17 | 4.714<br>15(T)* | 4.725<br>11 | 7. Report student progress to students and parents. | 4.324<br>17 | 3.600<br>22 | 4.172<br>15 | 18 |
| 20.3 | 4.124<br>22 | 4.286<br>21 | 4.429<br>19 | 8. Design and/or incorporate materials, illustrations and examples into lessons that blend with the cultural environments of students. | 4.209<br>20 | 3.833<br>18 | 40.20<br>19 | 19 |
| 3.7 | 4.832<br>4 | 4.954<br>3(T)* | 4.827<br>4 | 9. Use a variety of teaching skills (explains, questions, reinforces, etc.) | 4.565<br>5 | 4.143<br>10 | 4.239<br>9 | 8 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | Stem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 9.7 | 4.701<br>9 | 4.773<br>10(T)* | 4.733<br>10 | 10. Use a variety of instructional strategies (gaming, simulations, lecture, small group, individualized instruction, e.g.) | 4.393<br>15 | 3.950<br>14 | 4.185<br>12 | 13.7 |
| 3 | 4.878<br>2 | 4.909<br>6(T)* | 4.890<br>1 | 11. Communicate with students. | 4.684<br>2 | 4.454<br>4 | 4.330<br>3 | 3 |
| 15.3 | 4.536<br>15 | 4.714<br>15(T)* | 4.607<br>16 | 12. Communicate with other teachers. | 4.514<br>7 | 4.474<br>3 | 4.183<br>13 | 7.7 |
| 12 | 4.612<br>12 | 4.773<br>10(T)* | 4.611<br>14(T)* | 13. Communicate with supervisors. | 4.438<br>10 | 4.546<br>2 | 4.317<br>4 | 5.3 |
| 11.7 | 4.543<br>14 | 4.762<br>12 | 4.738<br>9 | 14. Communicate with parents. | 4.308<br>18 | 4.000<br>12(T)* | 3.819<br>21 | 17 |
| 21 | 4.310<br>21 | 4.273<br>22 | 4.407<br>20 | 15. Operate and use when appropriate audio-visual aids (e.g., projectors, recorders). | 4.369<br>16 | 3.857<br>17 | 4.253<br>6 | 13 |
| 4.7 | 4.801<br>5 | 5.00<br>1(T)* | 4.751<br>8 | 16. When necessary, changing plans based on the feedback received from students. | 4.478<br>9 | 3.864<br>16 | 4.182<br>14 | 13 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | Item | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 11 | 4.757<br>8 | 4.727<br>13(T)* | 4.684<br>12 | 17. Use a variety of classroom management (discipline) strategies. | 4.246<br>19 | 3.737<br>10 | 3.785<br>22 | 20 |
| 9 | 4.695<br>11 | 4.857<br>9 | 4.762<br>7 | 18. React to the needs of students parents, supervisors, other teachers and myself. | 4.403<br>13(T)* | 4.190<br>8 | 4.155<br>16 | 12.3 |
| 4 | 4.842<br>3 | 4.909<br>6(T)* | 4.874<br>3 | 19. Demonstrate openness and flexibility in teaching. | 4.640<br>4 | 4.273<br>6 | 4.265<br>5 | 5 |
| 4.7 | 4.795<br>6 | 4.954<br>3(T)* | 4.825<br>5 | 20. Exhibit a willingness to change when presented with valid data. | 4.652<br>3 | 4.409<br>5 | 4.344<br>2 | 3.3 |
| 1.3 | 4.885<br>1 | 5.000<br>1(T)* | 4.881<br>2 | 21. Demonstrate a willingness to continually improve as a teacher. | 4.787<br>1 | 4.591<br>1 | 4.547<br>1 | 1 |
| 6 | 4.782<br>7 | 4.952<br>5 | 4.801<br>6 | 22. Demonstrate adequate content knowledge. | 4.527<br>6 | 4.000<br>12(T)* | 4.207<br>10 | 9.3 |

*(T) = tie

student teacher performance items were ranked, giving a ranked order statement of student teacher performance. The highest rated SBTE rating of student teachers was on Item 21 (Willingness to Continually Improve as a Teacher). The lowest rating was on Item 2 (Apply Knowledge of Learning in Planning). Thus, for the SBTE ratings, Item 21 would be ranked 1st; Item 2 would be ranked 22nd.

Perceived Importance. While all items were rated quite highly by all groups, the highest three ranked items were:

No. 21. Demonstrate a willingness to continually improve as a teacher

No. 11. Communicates with students

No. 9. Use a variety of teaching skills.

The three lowest ranked items were:

No. 3. Writes clearly stated lesson and unit objectives

No. 8. Design/incorporate materials, etc., that blend with multi-cultural environment of students

No. 15. Operate and use A-V aids.

Student Teacher Performance. There is a much higher variance of raters' opinions of student teacher success on each of the items. However, the three items on which student teachers received highest ratings are:

No. 21. Demonstrate a willingness to continually improve as a teacher

No. 11. Communicate with students

No. 20. Exhibit a willingness to change when presented with valid data.

Five items appear to be about equal with respect to the lowest student ratings. These are:

No. 8. Design/incorporate materials, etc., that blend with multi-cultural environment of students

No. 6. Construct and use tests and other instruments to monitor student progress

No. 1. Apply knowledge of human growth and development in planning for instruction

No. 2. In planning for instruction apply knowledge about how students learn

No. 17. Use a variety of classroom management (discipline) strategies.

In looking at the ranked elements there are some interesting points:

1. In several instances, PTPP students excell at those competencies considered most important by the raters (communication with students, willingness to change).
2. In one case, students appear to perform poorly in the design and incorporation of multicultural elements. This was also a low ranking competency statement.
3. The element ranked lowest in terms of student performance (classroom management) is the element that appears in other success studies.[5] It appears that this is almost a nationwide concern based on published reports and informal sharing of teacher educators.
4. There is no indication in the data where the students feel they gained in competence. It is tempting to indicate that the success is a demonstration of programmatic impact. However, alternative explanations for a number of elements must also be considered. Given the evidence of a high level of interpersonal skills, it may well be that students gain competence in both program and extra-program experiences. Regardless, there is some high degree of success indication provided in the data.

[5]Fred L. Pigge, "Teacher Competencies: Need, Proficiency, and Where Proficiency was Developed," Journal of Teacher Education 29, No. 4, July-August, 1978.

Evaluation of the Evaluation

Generally, we are pleased with the kinds of data we collected on this first attempt. Our major problem was the long time delay between when the questionnaire data were collected - Spring 1977 - and when they were analyzed - Winter 1977. The reasons for this delay related to the coding of data, key punching, computer programming, and computer operations. The ultimate criterion of our effectiveness is whether the data were used and whether programmatic changes called for by the data actually occurred. On both accounts the answer was "Yes."

# FOLLOWUP STUDIES: PROVIDING THE THRUST FOR PROGRAM IMPROVEMENT*

Jerry B. Ayers
College of Education
Tennessee Technological University
Cookeville, Tennessee 38501

The problem of the evaluation of graduates of teacher education programs has been a concern of institutions of higher education for a number of years. Traditionally, evaluation efforts have consisted primarily of mailed questionnaires which solicited responses from graduates and/or supervisory personnel regarding effectiveness of the teacher education program. In turn, the results of these surveys were used for some program modifications and improvements in teacher education.

Historically, systematic evaluation of the graduates of the teacher education programs of Tennessee Technological University dates back to 1964 when a mail followup study was conducted of the graduates for the period 1958-1963. Prior to this time the evaluation efforts of the institution had been limited to information gathered about the performance of graduates through informal contacts with the college faculty and through contacts with principals, supervisors, and superintendents. From the period 1964 through 1969, the College of Education conducted yearly followup surveys of the graduates of the teacher education programs through the use of mail questionnaires.

The need for institutions of higher education to evaluate their teacher education programs was given emphasis when the 1970 standards for accreditation of teacher education were adopted by the National Council for Accreditation of Teacher Education (NCATE). Evaluation and the use of the results for program improvement were given further emphasis in the 1979 standards. Standard 6.1 states:

> The institution keeps abreast of emerging evaluation techniques and engages in systematic efforts to evaluate the quality of its graduates upon completion of their program of study and after they enter the teaching profession. This evaluation includes evidence of their performance in relation to program objectives.[1]

Standard 6.2 gives emphasis to use of evaluation results and states:

> The institution regularly evaluates its teacher education programs and uses the results of its evaluation in the modification and improvement of those programs.[2]

---

*Presented in part before the Annual Meeting of the American Educational Research Association, San Francisco, California, April, 1979.

In 1970, in order to meet the NCATE standards, gather curriculum evaluation data and sufficient information for curriculum development and reform, the College of Education initiated a series of systematic studies of the graduates of the teacher education programs. The general objectives of the evaluation program at Tennessee Technological University include the following:

1. To evaluate the objectives of the teacher education program of Tennessee Technological University through a systematic study of graduates of the program.

2. To provide information for faculty and administration concerned with teacher education programs at Tennessee Technological University in making decisions pertinent to curriculum evaluation and development.

3. To aid in the process of making long-range plans for improving the total program of the University with particular emphasis on teacher education.

## Types of Evaluation Studies

Three broad types of studies have been conducted at the University since 1970 as follows: (1) systematic surveys of the graduates of the teacher education programs and employers of the graduates, (2) special studies designed to provide specific information about a particular area of concern, and (3) implementation of the Tennessee Technological University Teacher Evaluation Model. Studies under groups 1 and 2 were begun in 1970 and use of the Evaluation Model was begun in 1973. Following is a brief account of each of the three types of studies.

### Surveys of Graduates and Employers

Each year, the graduates and the employers of the graduates for the past year are surveyed utilizing a specially designed survey instrument based on the objectives of the teacher education programs. This instrument allows the graduates to rate their level of attainment of the objectives of the teacher education programs and principals to rate what they feel is the level of attainment by the graduates of the objectives of the teacher education programs. These surveys aid in identifying where graduates are teaching and what problems they may have in their first year of teaching. Also, mail surveys have been made of graduates after they have been teaching as much as five years. Surveys of this nature have been used with both undergraduates and those individuals who completed programs at the Master of Arts and Educational Specialist levels.

It has been found that these surveys provide much information that can be used in program improvement and curriculum development. The surveys also provide a base for designing other studies and providing initial input for carrying out the Tennessee Technological University Teacher Evaluation Model.

## Special Studies

Throughout the past nine years, a variety of special studies have been conducted that have provided useful information toward meeting the NCATE standards and assisting with program development and redesign. Following is a partial list of studies that have been conducted. It is felt that the titles are self-explanatory with regard to use in meeting evaluation standards.

1. Predicting Quality Point Averages in Masters Degree Programs in Education
2. A Systematic Plan for the Restudy of a Teacher Education Program
3. Relationship of Student Grade Expectations, Selected Characteristics, and Academic Performance for Education Students
4. Attitudes of Students Toward Education Courses
5. Human Relationship Skills and Teacher Effectiveness
6. Study of the Graduates of the Counselor Education Program of Tennessee Technological University
7. Study of the Graduates of the Ed.S. Program of Tennessee Technological University for the Period 1971 Through 1974
8. Performance of Principals as Evaluated by Themselves, Superintendents, and Teachers
9. Investigation of the Relationships Between Authoritarianism of Principals and Their Ratings of Teachers
10. Study of the Graduates of the Library Science Program of Tennessee Technological University 1969-1973
11. Student Teachers Attitudes Toward Supervising Teachers
12. Longitudinal Study of Change in Teacher Dogmatism
13. Concurrent and Predictive Validity of the NTE

## Tennessee Technological University Teacher Evaluation Model

Even though the studies outlined above have provided valuable information, there was a need for the implementation of a more comprehensive model that would allow for collecting more direct information from the graduates of the teacher education programs. In a monograph prepared in 1970 for the American Association of Colleges for Teacher Education, Sandefur, suggested a model for evaluating the product of teacher education programs.[3] This model was based on generalizations made from the research literature on effective teaching and suggested procedures and instrumentation for conducting a product evaluation of teacher education programs.

In 1973, Tennessee Technological University developed and implemented a longitudinal study, based on Sandefur's general model, for systematic data gathering and for making decisions about its programs in teacher education. In 1978-79, the study entered its sixth year. The purpose of the study has been to adapt Sandefur's basic model for evaluation of teacher education graduates of the University and to apply the results of the study to the process of making long-range plans for improving the total educational program of the University with particular emphasis on the teacher education programs.

The specific objectives to be accomplished each year as a part of the study have been as follows:

1. To continue studying in a longitudinal manner those subjects who had previously participated in the application of the model.

2. To provide a descriptive profile of a sample of the previous years' graduates of the teacher education program of the University.

3. To determine relationships among selected variables as a measure of the total study.

4. To provide comparisons between the graduates of the teacher education programs of the University with those who might be considered effective teachers as defined in the original literature on teacher education.

5. To provide effective dissemination of research data to the faculty and the administration of the University associated with the teacher education programs.

6. To provide information and suggestions for curriculum evaluation and development based on empirical research.

7. Continue to evaluate the procedures employed in the study and to make long-range plans for modifications and refinements of the basic studies.

Sandefur's model and that developed by Tennessee Technological University differ in several respects. Both models emphasize the gathering of data from four sources including the graduate, supervisors, students of the graduate, and by independent observers. Instrumentation varies between the two models as does the entry point of the graduates into the evaluation process. Implementation of Sandefur's model at Western Kentucky University calls for the collection of initial data on subjects while they are student teaching. The subjects are then evaluated during their first, third and fifth year of teaching. Subjects of the Tennessee Technological University model enter during their first year of teaching and are evaluated each year for up to five years. Data analyses and uses of the data have been different at both institutions. The following paragraphs provide a more detailed explanation of the implementation of Tech's model.

In the fall of each year, all graduates for the past twelve months have been sent a questionnaire requesting routine demographic information in regard to their current employment, address, etc. Also each graduate has been asked to rate his attainment of the major objectives of the teacher education program and to evaluate certain key courses that he completed. Questionnaires were sent to all individuals who received the B.S. and M.A. In turn, all individuals who were teaching in Tennessee within a 100-mile radius of the University were asked to participate in the longitudinal followup study (about 70% of all graduates who were teaching were

within the defined limits of the study). Each year from 40-55 individuals have volunteered to participate in the study. In turn, these subjects have been followed throughout the course of the project. During 1977-78 the first group of subjects completed five years in the study. After 1977-78 it is anticipated that no further data will be collected on this first group. During 1977-78 a total of 129 graduates were participating in the study.

Data for the study have been collected each year by means of mail surveys, interviews, and observations in the classroom of the graduates. All subjects were visited each year on at least one occasion by a specially trained graduate assistant. The graduate assistant has observed in a subject's classroom for approximately a half day and recorded six to eight 20-minute periods of observation using a ten category system of interaction analysis. At the completion of all observation, the Classroom Observation Record and the Tuckman Teacher Feedback Form were completed.

The appropriate version of the Student Evaluation of Teaching (SET) was administered during each visit. The instrument has been completed by at least one class of students of each subject each year. While the students complete the SET, subjects who were participating in the project for the first year, completed the California F-Scale.

During the course of the day, the observer has interviewed each subject with regard to his opinions and ideas about the teacher preparation program of the University. Also the observers have asked each principal to complete the Principal's Questionnaire (an instrument which allows the principal to rate the level of attainment of the objectives of the teacher preparation program by the graduate) and the Teacher Evaluation by Supervisor Form.

Pertinent data such as quality point average, National Teacher Examinations scores, and other standardized measures have been gathered from permanent records. Figure 1 shows a summary of the sources of data, instrumentation, and use of data.

Data analyses have been confined largely to descriptive statistics and to the application of the t-test and analysis of variance to determine differences across years and correlational analyses to determine the relationships among the variables. Factor analysis techniques have been employed on a limited basis with certain data sets. It should be kept in mind that the data have been collected each year on as many as 150 variables.

Results of the study have indicated that the graduates who have been followed for as much as five years have changed little during the period. Few significant differences have emerged from the data. Therefore, it could possibly be concluded that a teacher's style may be developed early in his career in the classroom. Also, there have been few differences noted when comparing each group of first year

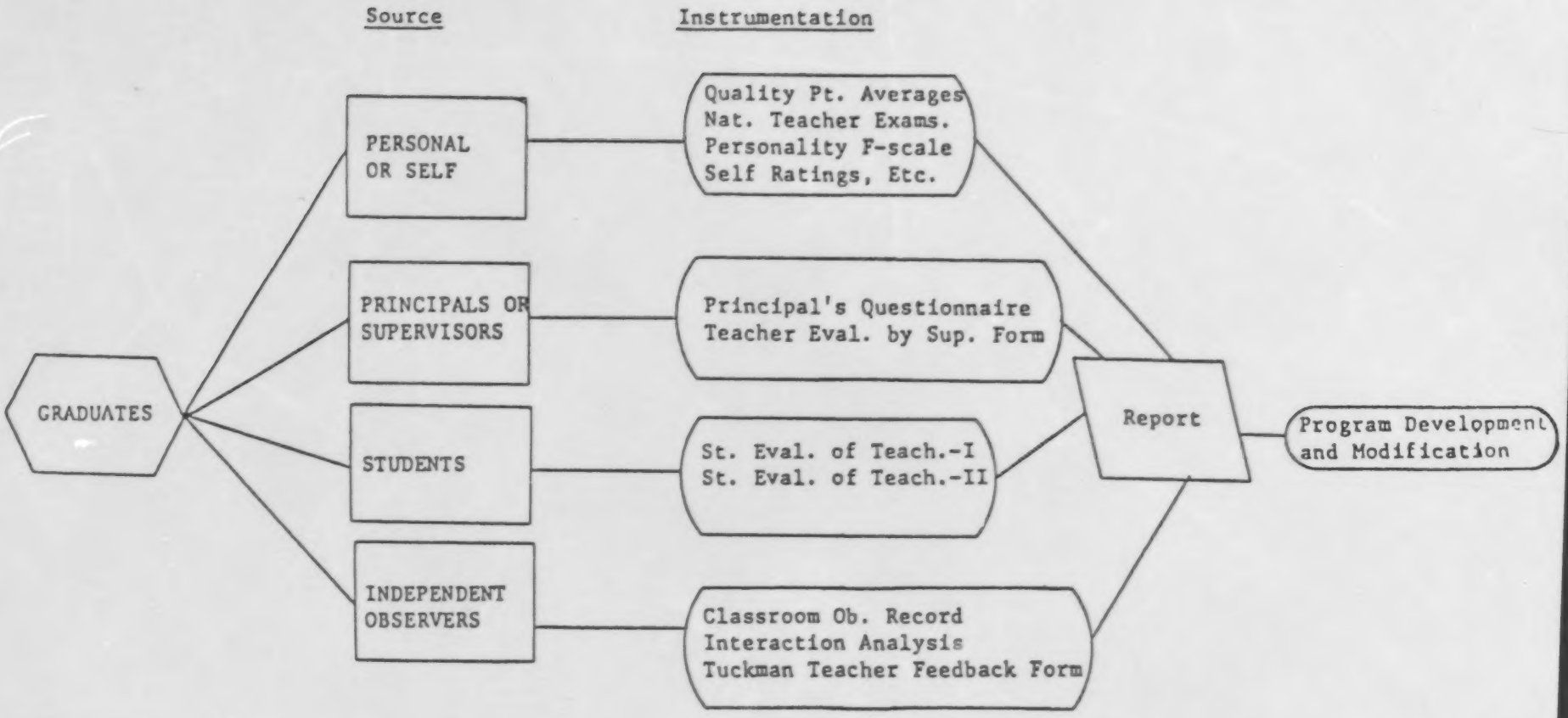



Figure 1. Summary of Sources of Data, Instrumentation and Use of Data.

graduates across the five years of the study. When comparisons of the data were made with that reported in the literature, the graduates appeared to be at a level comparable to or above the 50th percentile and were exhibiting many of the characteristics of good teachers. It appeared that two of the best predictors of success of teaching were level of achievement on the Professional Education Test of the NTE and overall grade point average in professional education and psychology courses. Less authoritarian graduates appeared to be overall better classroom teachers. The graduates appeared to have deficiencies in such areas as classroom control, cognitive knowledge of the sciences, insight into the characteristics of behavior and effective use of community resources. Limited factor analysis of the data gathered as a part of the use of the Classroom Observation Record indicated a factor pattern similar to that obtained by Ryans' in previous studies.

## Use of Evaluation Results

Program development at Tennessee Technological University is a continuous and ongoing process within the teacher preparation program. Program development, operation and evaluation at the University can be best described as a cyclic process that is continuous. Figure 2 shows a flow chart of the major activities involved in these operations. Key elements in the total process for program improvement are program monitoring (Activity 6) and evaluation of graduates (Activity 7). The former involves appropriate formative evaluation techniques while the latter utilizes the techniques discussed in the section of this paper entitled Types of Evaluation Studies. The remainder of the section describes some of the program improvement and modifications that have been made as a result of the evaluation studies.

Reports of all studies have been disseminated to all faculty in teacher education, members of the Teacher Education Committee, University administrators and other interested individuals. Meetings have been held for the purpose of discussing the results of these studies, with the departments involved and with other groups. The purpose of this wide dissemination of the reports has been to make the information readily available to those that are involved in developing, implementing and improving the programs of teacher preparation. A special curriculum study has been formed from the faculty within the teacher preparation programs of the University. This group has been charged with the responsibility of examining the results of the studies in depth. Following is a listing of some of the program changes and improvements at both the undergraduate and graduate levels that have been as a result of the studies.

### Undergraduate Level

1. Addition of more practicum and laboratory experiences into the teacher preparation program of undergraduates. Included under this broad area was more contacts in working with children in the schools.

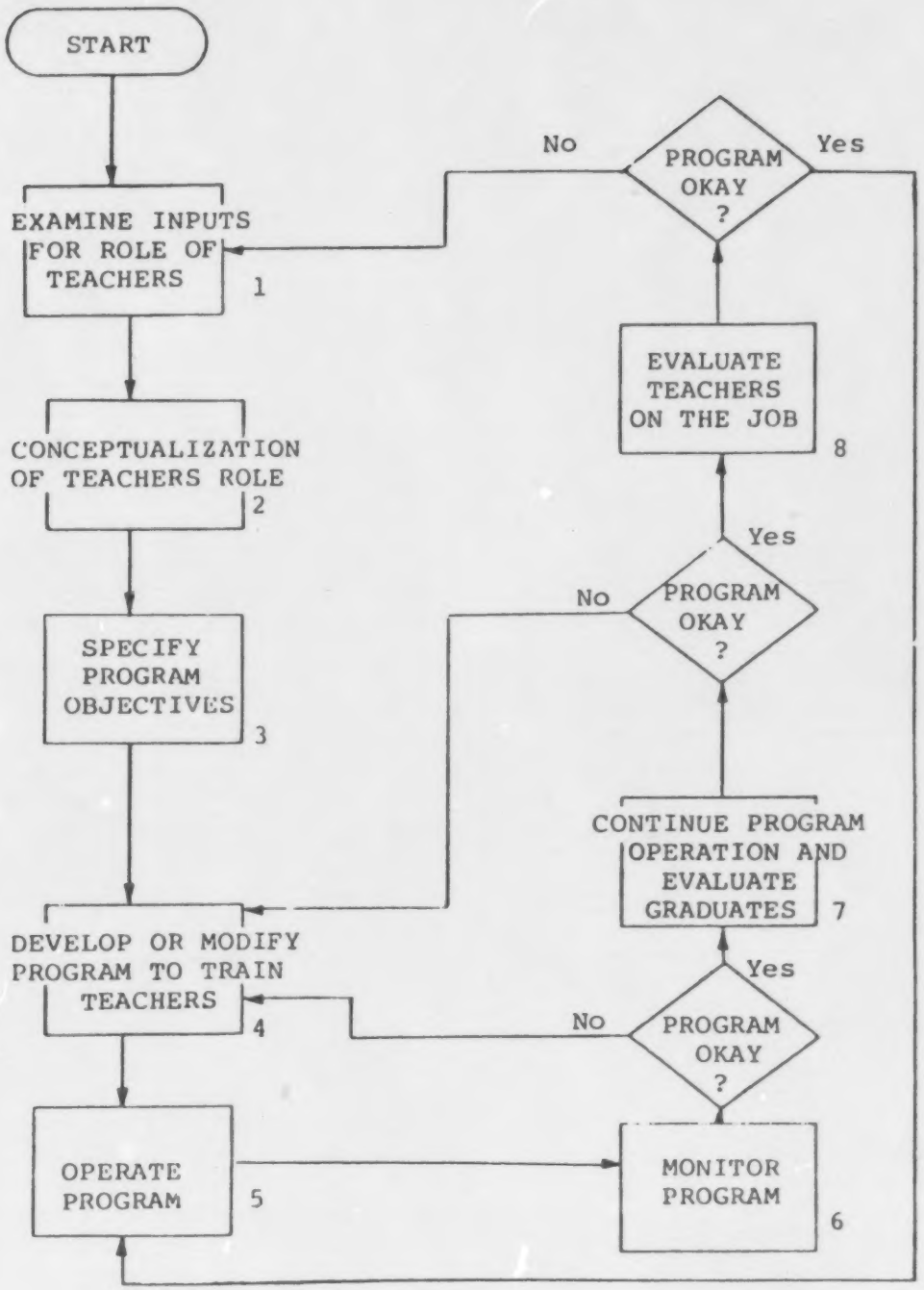



Figure 2. Flow Chart of Major Activities Involved in Program Development, Operation and Evaluation.

2. The course, Introduction to the Teaching Profession, (a freshman level course) was reorganized and made more practical and less theoretical.
3. Courses in the history, philosophy and social foundations of education were modified.
4. Overlap and duplication among courses has been eliminated.
5. Additional courses were created to meet special needs of various groups.
6. A science methods course for elementary teachers was created.
7. Reconceptualization of the teacher's role is taking place on a continuing basis.
8. Modifications in the requirements for a major in elementary education have taken place and the total program has been made more field oriented.
9. Secondary education major fields have been redesigned to include broad comprehensive teaching fields and sociology and geography were eliminated as separate teaching fields.
10. The major in health and physical education was divided into separate fields of study, i.e., a major in health or a major in physical education.
11. Course work in the Department of Educational Psychology and Counselor Education is being redesigned and placed on a competency basis.
12. Additional human relations training has been introduced into the curriculum.
13. A course has been developed in the area of special education that is now required for all students prior to graduation.
14. Special topics courses have been created in such areas as school law, discipline, and classroom interaction.
15. A special committee has been formed to study the total course offerings in the area of educational foundations.
16. Methods instructors have become more active in the public schools through actually assuming the role of the classroom teacher.
17. Efforts have been made to provide greater assistance to graduates in locating employment and in providing assistance after entry into the profession.

Graduate Level

1. The field experience courses of several of the programs of study were modified to make them more meaningful and of value to individual students.
2. Additional laboratory and field experiences were introduced in certain curricula.
3. Evaluative data have been used by the Department of Educational Psychology and Counselor Education in the design of a competency based program of study for prospective counselors and school psychologists.
4. Requirements for admission and the use of certain screening devices were modified.
5. Specific courses were reorganized and made more practical by relating theory to practice.

6. The research methods course was altered.
7. The core curriculum of the advanced programs was studied and modified.
8. The course in educational statistics is now more closely associated with the research course.
9. Library resources were improved and made more available to students taking courses off campus.
10. The courses in curriculum were re-evaluated and components were expanded to relate to individual needs of students.

## Summary

In summary, evaluation is an integral part of the teacher preparation program of Tennessee Technological University. A systematic plan has been developed and implemented for collecting information, performing necessary analyses, and dissemination of results. In turn, the evaluative efforts have been used in the cyclic process for program modification, improvement and development. Systematic evaluation is essential to the total process for the training of teachers and should be an integral part of every such process. However, change is a slow process and means need to be developed to expedite the process.

## Plans for the Future

It is anticipated that the study will be continued on an indefinite basis and that the trends that are beginning to emerge can be followed on a long-term basis. In turn, it is felt that the problems associated with teacher evaluation and implementation of the recommended standards for teacher accreditation are being met. Through the use of more sophisticated statistical techniques, additonal data analyses are planned. A continual plan of monitoring the literature has been implemented and it is anticipated that different instrumentation may be introduced into the study in the near future. Special studies utilizing data from the longitudinal model will be implemented. Plans are also underway to develop alternative models that could be used at institutions with lesser reserves.

## References

[1]Standards for Accreditation of Teacher Education. (Washington: The National Council for Accreditation of Teacher Education, 1977), p. 11.

[2]Ibid.

[3]Sandefur, J. T. An Illustrated Model for the Evaluation of Teacher Education Graduates. (Washington: American Association of Colleges for Teacher Education, 1970).

## Bibliography of Selected References to Model

Ayers, Jerry B. Report 74-4. Tennessee Technological University Teacher Evaluation Model. Cookeville, TN: Tennessee Technological University, College of Education, 1974, (mimeo.), 92 pp. ERIC No. 095 169.

Ayers, Jerry B. Report 75-4. Tennessee Technological University Teacher Evaluation Model-Year II. Cookeville, TN: Tennessee Technological University, College of Education, 1975. (mimeo.), 73 pp. ERIC No. 123 210.

Ayers, Jerry B. Report 75-5. Application of Tennessee Technological University Teacher Evaluation Model in Methods Classes in the College of Education. Cookeville, TN: Tennessee Technological University, College of Education, 1975. (mimeo.), 26 pp.

Ayers, Jerry B. Report 76-1. Tennessee Technological University Teacher Evaluation Model-Year III, Cookeville, TN: Tennessee Technological University, College of Education, 1976 (mimeo.), 92 pp. ERIC No. 126 085.

Ayers, Jerry B. Report 77-2. Tennessee Technological University Teacher Evaluation Model-Year IV, Cookeville, TN: Tennessee Technological University, College of Education, 1977 (mimeo.), 100 pp. ERIC No. 148 778.

Ayers, Jerry B. Report 78-2. Tennessee Technological University Teacher Evaluation Model-Year V. Cookeville, TN: Tennessee Technological University, College of Education, 1978 (mimeo.), 54 pp.

Ayers, Jerry B. "Followup Studies at Tennessee Technological University," Published in Forces for Molding Management and Control of Teacher Education (proceedings of the Sixth Annual Teacher Education Conference), Atlanta: Southern Regional Education Board, 1976, pp. 7-13. ERIC No. 152 695.

Ayers, Jerry B. "Implementation of a Longitudinal Model for Teacher Evaluation," Education, 96(3):218-21, Spring, 1976.

Ayers, Jerry B. and Merton J. Turck, "Longitudinal Study of Change in Teacher Dogmatism," College Student Journal, 10(1):84-7, 1976.

Dotson, James R. and Jerry B. Ayers. "A Systematic Plan for the Restudy of a Teacher Education Program," The Tennessee Tech Journal, 7:85-89, 1972.

# Critical Findings from the Teacher Preparation Evaluation Program: Teacher Entry, Retention, and Change

Presented by

Ronald D. Adams
Western Kentucky University

Annual Meeting of the
American Educational Research Association

San Francisco, California

April 11, 1979

The cost of printing this publication by Western Kentucky University was paid from state funds KRS 57.375.

## Introduction

If the standards for accreditation of teacher education programs are examined, it becomes evident that the organized professional groups responsible for adopting those standards consider the systematic evaluation of teacher education important, even essential, to the improvement of teacher education efforts and ultimately to public education in this country. NCATE and NASDTEC, the two organizations most concerned with specific accreditation of teacher education programs, have standards and recommendations which emphasize evaluation of programs through the systematic follow-up of teachers both while in the undergraduate programs and after they enter the profession.* Yet there exist very few examples of evaluation programs that attempt systematic, continuous evaluation of teacher education efforts. Pegues (1978) after conducting a national survey of teacher education institutions concluded, "There is only moderate contemporary involvement in systematic follow-up evaluative practices by teacher education programs . . . . There is a minimal level of evaluative sophistication in the design and the implementation of contemporary evaluation schema."

This report is a brief discussion of one evaluation program that has made a serious attempt to meet the intent and spirit of the evaluation charge issued through Standard 6.1 of the NCATE standards (1977) and Standard 3.1 of the NASDTEC recommendations (1976). The Western Kentucky University Teacher Preparation Evaluation Program (TPEP) represents the eight year commitment of an institution to the evaluation of teacher education graduates and has provided a rare set of longitudinal data not often found in educational research literature. Thus, it is important to consider the source of the reported findings and nature of the program that produced them.

---

*NCATE - National Council for Accreditation of Teacher Education

NASDTEC - National Association of State Directors of Teacher Education and Certification

Background

In 1972 J. T. Sandefur (1972) authored a monograph entitled, An Illustrated Model for the Evaluation of Teacher Education Graduates that served as the basis for the Teacher Preparation Evaluation Program (TPEP). Sandefur's model was founded on research drawn from teacher effectiveness studies and advocated a systematic, longitudinal approach to the evaluation of teacher education graduates. While certain specific instruments and procedures for data collection were suggested by Sandefur, the model stressed three general concepts which were different from the more common practices of follow-up mailed questionnaires. These concepts were 1) that data be objective and quantifiable, 2) that data be collected from multiple sources to include direct classroom observation by trained observers, and 3) that graduates be followed for a minimum of five years with data collected during specified points during that period.

In the spring of 1972, TPEP was initiated with data collected from a sample of 40 student teachers. Since that modest beginning, data will have been collected from 307 student teachers, 170 first year teachers, 76 third year teachers, and 38 fifth year teachers by the end of the 1979 spring semester.

Procedures

TPEP is organized around a Cycle-Phase arrangement in which a new sample is selected each year to begin a new Cycle. Each Cycle has four Phases that correspond to data collection periods. Phase 1 represents data collection during student teaching, with Phases 2, 3, and 4 corresponding to data collection during the first, third, and fifth years of teaching experience, respectively. Figure 1 displays this Cycle-Phase arrangement. Presently data are being collected from Cycle III Phase 4, Cycle V Phase 3, Cycle VII Phase 2 and Cycle VIII Phase 1.

| | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 |
|---|---|---|---|---|---|---|---|---|
| CYCLE I | Phase 1 | Phase 2 | | Phase 3 | | Phase 4 | | |
| CYCLE II | | Phase 1 | Phase 2 | | Phase 3 | | Phase 4 | |
| CYCLE III | | | Phase 1 | Phase 2 | | Phase 3 | | Phase 4 |
| CYCLE IV | | | | Phase 1 | Phase 2 | | Phase 3 | |
| CYCLE V | | | | | Phase 1 | Phase 2 | | Phase 3 |
| CYCLE VI | | | | | | Phase 1 | Phase 2 | |
| CYCLE VII | | | | | | | Phase 1 | Phase 2 |
| CYCLE VIII | | | | | | | | Phase 1 |

Figure 1 - Cycle and Phase Arrangement

Data have been collected from multiple sources to include 1) direct classroom observation by trained observers, 2) peer and supervisor ratings, 3) student evaluations, 4) university records, and 5) participant self reports. Instruments were chosen that had demonstrated research merit, that could be administered with reasonable ease, and that contributed to the overall objectives of the program. The longitudinal nature of the project has prohibited frequent change in instruments even when new research findings offered tempting avenues for change. However, selected instruments have been added and existing instruments modified as the need and opportunity arose. Thus, data are not complete on all instruments for all Cycles and Phases. Table 1 presents the TPEP instrumentation and sources of data. Brief descriptions of these instruments are as follows:

Career Base Line Data

Career base line data for participants are obtained primarily from three sources--the Career Base Line Data Questionnaire (CBLD), the Teacher Preparation Evaluation Inventory (TPEI), and transcripts of grades. The CBLD questionnaire is designed to obtain demographic data from participants not readily available from other sources. Initial personal and professional background information is collected for student teachers and updated as participants continue in the program.

Table 1

Categories of Data Collection for TPEP

| Variable | Source | Instrument |
|---|---|---|
| Career Line Data | Teacher | Demographic data questionnaire<br>Undergraduate GPA's<br>Teacher Preparation Evaluation Inventory<br>Follow-up Questionnaire<br>Ratings of Teacher Education Programs |
| Standardized Measures | Teacher | F-Scale (authoritarianism)<br>Rokeach Dogmatism Scale<br>Teacher Concerns Checklist |
| Direct Classroom Observation | Trained Observers | Interaction Analyses<br>Classroom Observation Record<br>Observer Checklist |
| Pupil, Peer, and Supervisor Evaluations | Pupils<br>Peers<br>Supervisor | Student Evaluation of Teaching I and II<br>Teacher Evaluation by Peer/Supervisor |

The perceptions of participants toward the university's preparation program is obtained through administering the Teacher Preparation Evaluation Inventory. In addition, a measure of the problems encountered in teaching are obtained by this instrument. Participants are asked to respond to a Likert-type scale for a number of items as well as to four open-ended items.

Four grade point averages (GPA) are computed from the participants' transcripts of grades. Overall GPA, professional education GPA, subject GPA, and student teaching grade are entered into the data files as separate variables at the completion of the student's undergraduate teacher preparation program.

Direct Classroom Observation

Two direct classroom observation systems are employed to obtain data on participants' classroom behavior. Observers are trained in the use of these systems, and interobserver reliability is established each year. The Classroom Observation Record, developed by Ryans (1960), is used to assess four dimensions of pupil behavior and eighteen dimensions of teacher behavior. Each dimension is carefully described and defined in a glossary accompanying the recording form. A seven interval scale is used to rate each of the pupil and teacher dimensions immediately after each observation period. Shiek and Adams (1978)

found that sixteen of the eighteen teacher behaviors formed three major factors.* Factor A can be described as organized, confident, and systematic classroom behavior; Factor B as empathetic, understanding, and adaptable classroom behavior; and Factor D as stimulating, original, and alert teaching behavior. A fourth factor, Factor C, is composed of ratings of student behaviors characterized by alert, responsible, and confident classroom behavior. Two additional teacher dimensions not included in the factor structure are "fairness" and "attractiveness".

Interaction Analysis (IA)

A nineteen category Classroom Interaction Analysis System provides data on teacher-student verbal interaction and, to a lesser extent, nonverbal behavior. This system is a modification of the Flanders' (1970) system of interaction analysis by Adams and Sandefur (1976). Twelve categories of teacher talk, three categories of student talk, and four nonverbal categories comprise the system. The observer records a numerical value corresponding to a specific category every three seconds or every time the category changes. Thus, an objective record is obtained of verbal interaction occurring in the classroom. Two twenty minute observations per participant are recorded. Appendix C contains the specific categories and the variables obtained from the CIAS.

Pupil, Peer, and Supervisor Evaluations

Perceptions about each participant are obtained from pupils, peers, and supervisors. These data are collected for each phase of TPEP with the only exception occurring in Phase 1. For Phase 1 the cooperating teacher rather than peers and supervisors rates the participant as a student teacher. Peers and supervisors are asked to rate participants on two instruments, TEP/S and TEP/S II. The TEP/S instrument was developed at Kansas State Teachers' College and obtains ratings on four broad items related to teacher behavior. The TEP/S II instrument

*The COR factor structure reported here closely approximates that found by Ryans (1960).

was developed by Adams (1978) and consists of seventeen items designed to measure three areas of perceived teacher behavior. The TEP/S II has recently been developed and limited data are available at this time.

Perceptions of pupils about each participant are obtained from the SET I and II. SET I, developed by Veldman and Peck, (1970) allows students to rate the teacher on ten items that measures five dimensions of teaching behavior. This instrument was derived from the Pupil Observation Survey Report developed by McClain. Veldman (1970) found the SET I could be used with pupils in the third grade and above. However, use of this instrument has been limited to grade four and above for the TPEP study.

The Student Evaluation of Teaching II, developed by Haak, Kleiber, and Peck, (1970 measures three dimensions of teaching as perceived by pupils. True-False responses are obtained from 23 statements regarding the teacher's behavior and feelings toward students. Two forms of this instrument allow data to be collected from pupils in grades kindergarten through sixth. A verbal, non-pencil-paper, card form is administered to kindergarten through third grade students while a machine-scorable form is administered to fourth through sixth grade students.

Standardized Measures

Three instruments are administered to participants to obtain measures of authoritarianism, dogmatism, and level of concern about teaching. These data are collected for each Phase of TPEP.

The F-Scale, forms 45 and 40, was developed by Adorno and others (1964) to measure individual prejudices and antidemocratic tendencies. This 28 item scale refers to opinions regarding a number of social groups and issues.

To provide supplementary data regarding dogmatic tendencies, the Rokeach Dogmatism Scale (1960) is administered to each participant along with the F-Scale. This 40 item scale measures dogmatic tendencies by requiring the participant to respond to a six-point, agree to disagree, scale for each item. Since the Rokeach scale was added in 1975, data are incomplete for earlier participants.

A recent addition to the instrumentation of TPEP is the Teacher Concerns Checklist (George, 1974). This instrument allows for measures of the concerns teachers have about the teaching profession and students. The first administration of this instrument in TPEP was during the Spring 1976 data collection period.

Graduate assistants have been the primary source of manpower for data collection. The program director has conducted the training of graduate assistants in the two observation systems: the Classroom Observation Record and the Classroom Interaction Analysis System. A minimum interobserver reliability standard of .75 has been adhered to and consistency across years has been maintained by the director. In addition, training in the administration of the Student Evaluation of Teaching I and II has been given to each graduate assistant observer prior to data collection.

Actual data collection consist of two visits to each participating graduate toward the latter part of the spring semester every other year for five years. Classroom observations are made during each visit, and packets of instruments are left for the participant, two selected peer teachers and the supervisor, at the first visit. Peer teachers and the supervisor are requested to return their completed rating forms, the Teacher Evaluation by Peer/Supervisor I and II, via a self-addressed pre-postage paid envelope. The participants' packet, consisting of standardized measures of teacher characteristics and concerns and various demographic and teacher perceptions data, are collected at the second visit. Also, during the second visit, a student evaluation of the participating teacher is obtained.

A major change in TPEP occurred in 1977 and 1978 when the sampling procedures changed from the random selection of twenty elementary and twenty secondary student teachers to a larger initial sample of student teachers. This procedure change was due to the low rate of student teachers who enter

the profession after graduation and to the dropout rate of teachers after the first year of teaching. Present plans include obtaining selected self-report data on all student teachers and conducting a follow-up of a random sample of 40 who enter the teaching profession in Kentucky the first year after graduation. This procedure will maximize the number of participants during student teaching and during their first year of teaching. The loss of observation data and student evaluations for student teachers is compensated for by the increased numbers in Phase 1 and Phase 2. Table 2 presents the current status of participants.

TABLE 2

Number of Teachers Who Remained in the Study Sample Over a Six-Year Period

| Cycle | Level | Phase 1 | Phase 2 | Phase 3 | Phase 4 |
|---|---|---|---|---|---|
| | | Student Teaching | 1st Year Teaching | 3rd Year Teaching | 5th Year Teaching |
| Cycle I (1972) | Elementary | 20 | 15 | 10 | 6 |
| | Secondary | 18 | 7 | 3 | 3 |
| Cycle II (1973) | Elementary | 20 | 13 | 7 | 6 |
| | Secondary | 20 | 8 | 5 | 4 |
| Cycle III (1974) | Elementary | 20 | 16 | 13 | 11 |
| | Secondary | 17 | 9 | 8 | 8 |
| Cycle IV (1975) | Elementary | 17 | 14 | 11 | No data for Cycle IV Cycle V Cycle VI and Cycle VII |
| | Secondary | 20 | 6 | 4 | |
| Cycle V (1976) | Elementary | 20 | 12 | 8 | |
| | Secondary | 20 | 5 | 7 | |
| Cycle VI (1977) | Elementary | 39 | 21 | No data for Cycle VI and Cycle VII | |
| | Secondary | 32 | 12 | | |
| Cycle VII (1978) | Elementary | 26 | 21 | | |
| | Secondary | 18 | 11 | | |
| Total | Elementary | 162 | 112 | 49 | 23 |
| | Secondary | 145 | 58 | 27 | 15 |

Analyses

Data are stored in a "master" data set that allows flexible management of data and that is compatable with various statistical packages. New data are added to the master data set each year for both the repeated measures across Phases and data from new samples of participants. A second data set is maintained for storage and analyses of student teacher data.

The large and rather complex data set affords opportunities for multiple types of analyses on numerous problem areas. The continuing nature of TPEP also contributes to the continuous nature of data analyses. The analyses reported in this paper should be considered more as representative of the stage of development in TPEP analyses than as a definitive, comprehensive or completed work. Problems with data coding and processing have limited, somewhat, the intended presentation; however, a complete analyses on the topics of teacher attrition and change is beyond the scope of this paper. Thus, selected analyses and findings are offered that appear to provide some insight into these topics.

Problem One: Factors related to teachers' failure to enter and/or remain in teaching.

Each fall a survey of TPEP participants was conducted to determine their occupational status. If not teaching, they were asked to give reasons why. Table 3 presents reasons given for not entering or for leaving the teaching profession. A major limitation of the TPEP is the lack of resources to follow non-teaching graduates to determine if they enter teaching at a later time. Presently, if a participant does not enter teaching the next year after graduation, or if a teacher leaves teaching after employment they are eliminated from further follow-up.

From Table 3 the most often given reasons for not entering teaching is lack of an available or desirable teaching job, accounting for 42% of the total reasons. A small (8%) but disturbing percentage of teachers reported their reason for not entering teaching was that they did not want to teach; another 10% of first year

dropout teachers gave the same reason. This percentage was determined by only 10 teachers, but represents approximately eight percent of the total teachers who left teaching since 1972. Reasons for leaving teaching between the first and third years and the third and fifth years appear to be family oriented, i.e., marriage and/or pregnancy.

Table 3

Percent of TPEP Participants Who Have Not Entered Teaching During Years 1, 3, and 5 by Reasons Given

| Reasons | Attrition After Graduation | | | Attrition Between 1st and 3rd Year | | | Attrition Between 3rd and 5th Year | | |
|---|---|---|---|---|---|---|---|---|---|
| | Elem. (n=30) | Sec. (n=65) | Total | Elem. (n=19) | Sec. (n=11) | Total | Elem. (n=6) | Sec. (n=1) | Total |
| Salary | 0 | 5 | 3 | 0 | 9 | 3 | 0 | 0 | 0 |
| Graduate Student | 7 | 6 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| Marriage | 7 | 3 | 4 | 11 | 0 | 7 | 33 | 0 | 29 |
| Pregnancy | 3 | 0 | 1 | 26 | 9 | 20 | 17 | 0 | 14 |
| Military | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Did Not Want To Teach | 7 | 9 | 8 | 5 | 18 | 10 | 0 | 0 | 0 |
| No Job in Location | 17 | 18 | 18 | 16 | 18 | 20 | 0 | 0 | 0 |
| No Job Anywhere | 20 | 26 | 24 | 5 | 0 | 3 | 0 | 0 | 0 |
| Other | 30 | 22 | 26 | 37 | 36 | 37 | 50 | 100 | 57 |
| Data Missing | 10 | 12 | 12 | 0 | 9 | 3 | 0 | 0 | 0 |
| Total TPEP Participants | 145 | 123 | | 68 | 39 | | 31 | 15 | |

In addition to self-report reasons for not entering or remaining in teaching, analyses were made of selected variables to determine if factors could be identified from the TPEP data set that were related to the attrition of graduates. A discriminate analyses statistical technique was employed from the Statistical Package for the Social Sciences (1975) to determine a multiple correlation coefficient between selected sets of independent variables and membership in the teaching group or attrition group. Variables measured in student teaching were the independent variables used to predict entry or non-entry into teaching, while measures taken at the end of the first year of teaching were used to predict attrition between the first and third year of teaching.

In predicting attrition, often n-sizes were less than desirable due to attrition from student teaching into the teaching profession and/or due to the time an instrument was introduced into TPEP. In some cases the n-size and number of independent variables produced suspected spuriously high correlations. However, n-sizes are reported in such cases and the researcher may make his/her own interpretation of the analyses. When correlations reached .40 or higher,additional tables were constructed to report how well that particular set of variables was able to predict the outcomes of attrition. The first analyses will deal with attrition from student teaching into first year teaching followed by analyses of attrition from first year teaching to third year teaching.

Analyses of Attrition from Student Teaching to First Year Teaching

Selected variable sets were analyzed to determine if factors could be identified that related to attrition of graduates from entering teaching after graduation. Variables measured from multiple sources were selected for the analyses to obtain predictor variable sets from a broad perspective. Results of these analyses are presented in Table 4.

Low multiple correlation coefficients were found for each set of variables when corelated with attrition/teacher group membership. It would appear that the selected variables for these analyses were not factors related to entry into teaching.

Table 4

Multiple Correlation Coefficients of Selected Variables and Attrition/Teaching Group Membership from Student Teaching to First Year Teaching

| Variable Set & Source | Grade Level | Group | Group n | Variable n | Multiple Correlation |
|---|---|---|---|---|---|
| GPA | Sec. | Att. | 79 | 4 | .245 |
| | | Tch. | 43 | | |
| (Univ. records) | Elem. | Att. | 41 | 4 | .222 |
| | | Tch. | 87 | | |
| TCC | Sec. | Att. | 37 | 5 | .243 |
| | | Tch. | 16 | | |
| (Participant) | Elem. | Att. | 22 | 5 | .338 |
| | | Tch. | 31 | | |
| Demographic | Sec. | Att. | 82 | 2 | .075 |
| | | Tch. | 45 | | |
| (Participant) | Elem. | Att. | 42 | 2 | .094 |
| | | Tch. | 90 | | |
| COR Factor | Sec. | Att. | 81 | 6 | .087 |
| | | Tch. | 46 | | |
| (Observers) | Elem. | Att. | 42 | 6 | .286 |
| | | Tch. | 91 | | |
| SET I | Sec. | Att. | 38 | 3 | .079 |
| | | Tch. | 16 | | |
| (Students) | Elem. | Att. | 27 | 3 | .136 |
| | | Tch. | 46 | | |
| IEF/S | Sec. | Att. | 74 | 4 | .204 |
| | | Tch. | 45 | | |
| (Supervising teacher) | Elem. | Att. | 41 | 4 | .767 |
| | | Tch. | 86 | | |

Analyses of Attrition from First Year Teaching to Third Year Teaching

As for the previous analyses, variables were selected that represented measurement from multiple sources. The sample sizes for these analyses were understandably much smaller than for the entry attrition analyses. The nature of the multiple correlation technique utilized is such that as the number of variables approach the number of cases, the multiple correlation approaches 1.00. Thus, there is concern that some correlations that were above .40 may be spuriously high due to n-size and are not true indicators of the relationship between the predictor variable set and group membership. To assist somewhat in this problem, only the relationships above .40 that had at least a 5 (cases) to 1 (predictor variable) ratio were considered for further analyses. Even in these analyses, there is suspicion of inflated correlations. Table 5 contains the results of these analyses.

Table 5

Multiple Correlation Coefficients of Selected Variables and Attrition/Teaching Group Membership from First Year Teaching to Third Year Teaching

| Variable Set & Source | Grade Level | Group | Group n | Variable n | Multiple Correlation |
|---|---|---|---|---|---|
| TCC | Sec. | Att. | 6 | 5 | .678 |
| | | Tch. | 13 | | |
| (participant) | Elem.* | Att. | 15 | 5 | .410 |
| | | Tch. | 28 | | |
| COR Factors | Sec. | Att. | 16 | 6 | .317 |
| | | Tch. | 29 | | |
| (observers) | Elem. | Att. | 31 | 6 | .250 |
| | | Tch. | 59 | | |
| SET I | Sec. | Att. | 14 | 5 | .213 |
| | | Tch. | 25 | | |
| (students) | Elem. | Att. | 9 | 5 | .226 |
| | | Tch. | 20 | | |
| SET II | Sec. | Att. | 7 | 3 | .310 |
| | | Tch. | 12 | | |
| (students) | Elem. | Att. | 15 | 3 | .123 |
| | | Tch. | 30 | | |
| TEP/S | Sec. | Att. | 8 | 4 | .455 |
| | | Tch. | 9 | | |
| (peer) | Elem. | Att. | 14 | 4 | .308 |
| | | Tch. | 29 | | |
| TEP/S | Sec.* | Att. | 13 | 4 | .431 |
| | | Tch. | 24 | | |
| (supervisor) | Elem. | Att. | 26 | 4 | .159 |
| | | Tch. | 50 | | |
| F-Scale & Rokeach | Sec. | Att. | 7 | 2 | .180 |
| | | Tch. | 21 | | |
| (participant) | Elem. | Att. | 18 | 2 | .329 |
| | | Tch. | 39 | | |
| Interaction Analyses | Sec. | Att. | 16 | 10 | .595 |
| | | Tch. | 29 | | |
| (observer) | Elem. | Att. | 28 | 10 | .355 |
| | | Tch. | 59 | | |

* These variable sets met the minimum standards for further analyses.

These results are similar to those found for entry attrition. There does not appear to be a strong factor among the selected variable sets that is related to attrition/teaching group membership. The two variable sets of "teacher concerns" for elementary and "supervisor ratings" for secondary minimally met the established criteria, but were not impressive in the strength of relationship found. Table 6 and Table 7 present the predictive strength of these variable sets.

Table 6

Number and Percent of Attrition/Teacher Group Membership Cases Predicted by the Teacher Concerns Checklist Variable Set for Elementary TPEP Participants*

| | Correctly Predicted | | Incorrectly Predicted | |
|---|---|---|---|---|
| | n | % | n | % |
| Attrition Group | 8 | 53 | 7 | 47 |
| Teaching Group | 23 | 82 | 5 | 18 |
| Total Group | 31 | 72 | 12 | 28 |

* Multiple Correlation = .410
The Mean concerns were higher for the teaching group than the attrition group.

Table 7

Number and Percent of Attrition/Teacher Group Membership Cases Predicted by the Supervisor Ratings from the Teacher Evaluation by Peer/Supervisor for Secondary TPEP Participants*

| | Correctly Predicted | | Incorrectly Predicted | |
|---|---|---|---|---|
| | n | % | n | % |
| Attrition Group | 9 | 69 | 4 | 31 |
| Teaching Group | 20 | 83 | 4 | 17 |
| Total Group | 29 | 78 | 8 | 22 |

* Multiple Correlation = .431
Mean ratings were generally higher for the teacher group than the attrition group.

Summary

A possible explanation for the lack of correlation between attrition/teacher group membership and the sets of predictor variables could be the gross measures of teacher attrition after graduation. For example, graduates who do not go into teaching because of graduate school may be different from graduates who do not enter teaching because of a lack of available position. These analyses did not attempt to study the attrition group by reason for not entering teaching, but grouped all non-teaching graduates into one attrition group. If this factor was considered, stronger relationships may be found.

A second possible explanation for the lack of correlations found may be the type of analyses used for predicting group membership. An analysis of variance technique for individual variables may be more meaningful in identifying factors than the multiple correlation approach utilized in these analyses. Through casual observation of the data there appeared to be differences in the means of the attrition group and the teaching group though no tests of significance have been made.

Future expansion of TPEP as described earlier will permit more exploration of these and other variables through the increase of initial n-size (all student teachers) and the increase in the follow-up sample of first year teachers. For example, it should be possible, with larger samples, to study grade levels and subject areas as unique groups. The findings, while somewhat disappointing, suggest a more complex set of factors operating that are related to attrition, than were studied in this paper.

Problem Two: Change in teacher behavior over three and five years of teaching experience.

The repeated measures of variables from student teaching through five years of teaching experience offers a longitudinal data base for which analyses of change can readily be made. Longitudinal data are rare due to obvious continuing resource commitments by institutions. The ability of TPEP to provide such data offers a unique look at teachers as they mature in the profession. These analyses will focus on data collected through direct classroom observation and will utilize a two-way repeated measures analyses of variance technique from the Bio Medical (1975) statistical package. The analyses of variance technique tests for differences between means 1) across Phases or years of teaching, 2) across Cycles or samples, and 3) the interaction effects between Phase and Cycle.

The two sets of variables are from the Classroom Observation Record (COR) and the Classroom Interaction Analyses System (CIAS), a modification of the Flanders interaction analyses system. Each set of data is discussed separately.

The Classroom Observation Record (COR) consists of 22 dimensions of teacher and pupil behavior each of which is judged by a trained observer along a seven point scale. These 22 items have been reduced through factor analyses (Shiek and Adams, 1978) into four factors. These factors are as follows:

Factor A - Organized, confident, and systematic teacher behavior.

Factor B - Empathetic, understanding, and adaptable teacher behavior.

Factor C - Alert, responsible, and confident student behavior.

Factor D - Stimulating, original, and alert teacher behavior.

Two additional dimensions not included in the four factor solution were "fair" teacher behavior and teacher "attractiveness."

The first analyses were made of Phase 1, student teacher data, through Phase 3, third year teacher data. Four Cycles have been completed for both secondary and elementary participants, providing a reasonable n-size (n=39 for elementary, n=19 secondary) for these analyses. Figure 2 presents the outcomes. Source tables from which graphs were taken are contained in Appendix A.

Elementary teachers had a clear increase ($p < .05$) in desirable teaching behavior during the third year of teaching over student teaching and first year teaching behavior for Factors A, B, and C. The increase for Factor D was not as dramatic ($p < .10$) but maintained a pattern similar to the other Factors. A similar pattern was observed for secondary teachers for Factors A and D with increases significant beyond the .05 level. For Factor B, the increase appeared more linear with increases in desirable behavior occurring the first and third year over student teaching classroom behavior. While there seemed to be increases across years in Factor C, the difference was not significant at the .05 level.

When differences in means across Cycles were examined, no significant main effects were noted. Likewise, there were no significant effects observed for Cycle Phase interaction.

In the second set of analyses data were analyzed across a six year period from student teaching through the fifth year of teaching. Figure 3 presents these analyses. Two cycles have been completed affording an available n-size of 10 for elementary and 7 for secondary. While these n-sizes are small, there were significant results obtained beyond the .05 and/or .10 levels for all factors of the COR. As was the case in the previous analyses, main effects differences across cycles and interaction effects were not found to be significant at the .05 level or beyond. Appendix B contains tabular presentation of these analyses.

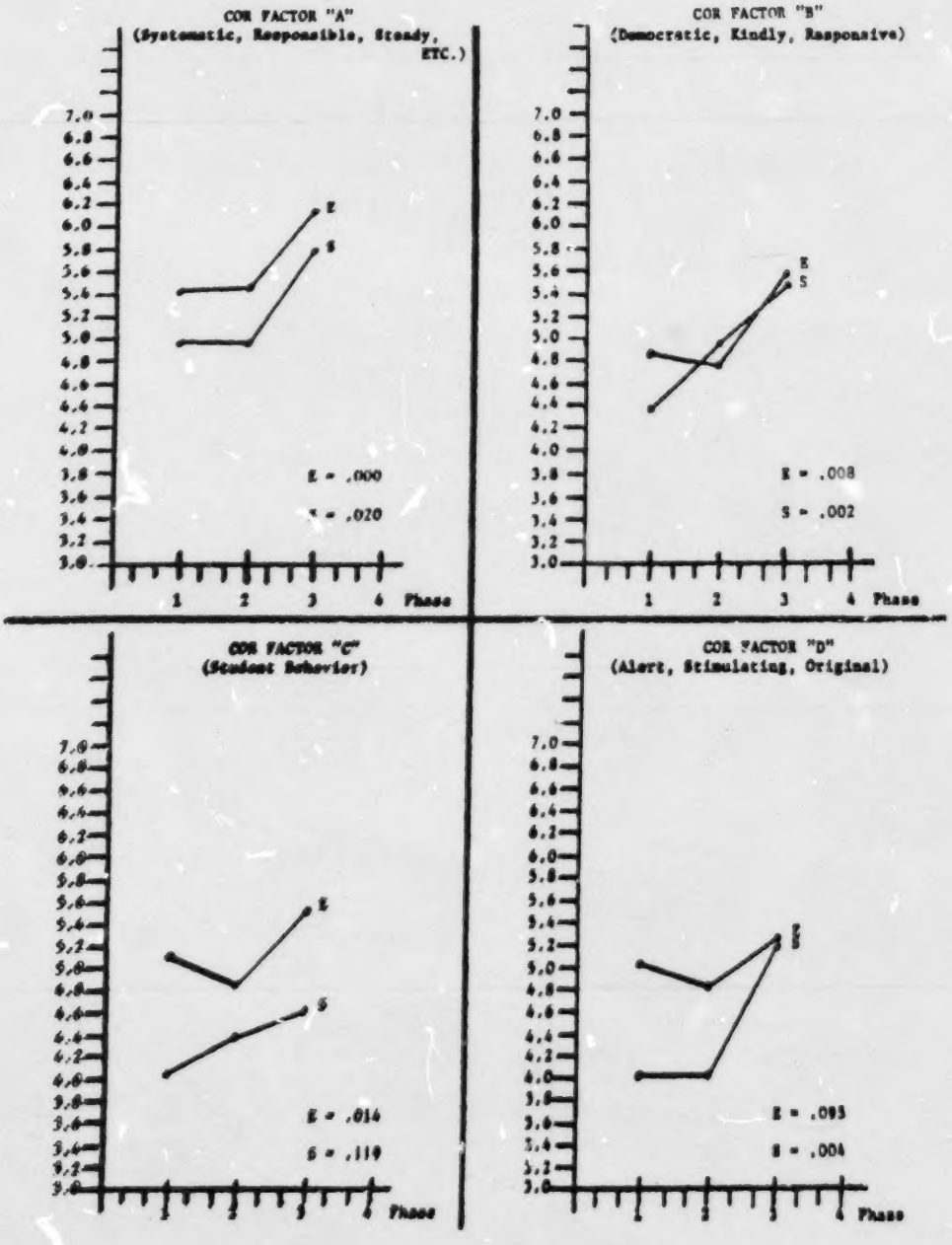



*Figure 2 - Graphs of COR Repeated Measures ANOVA across TPRP Phases 1, 2, and 3.*

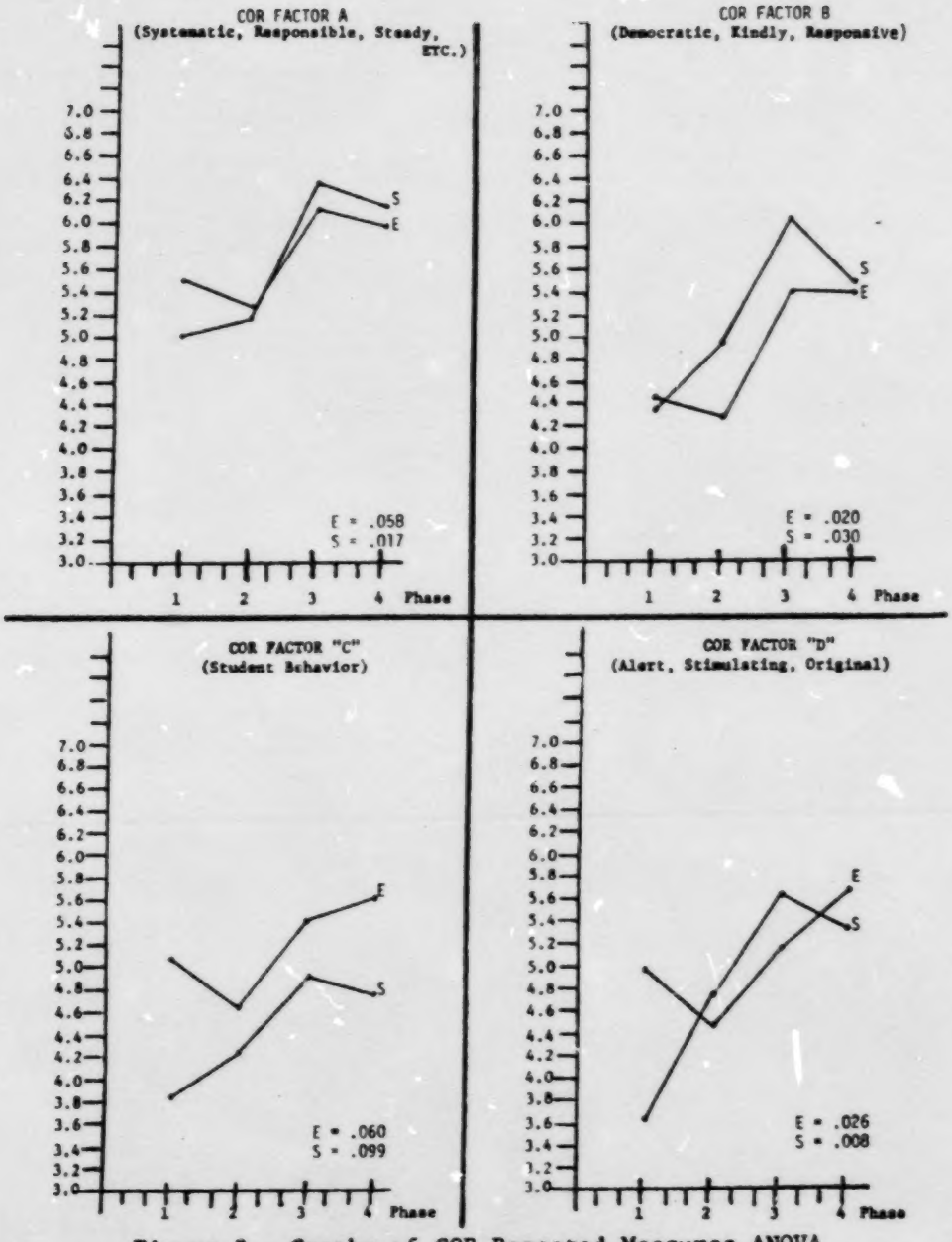



Figure 3 - Graphs of COR Repeated Measures ANOVA Across TPEP Phases 1, 2, 3, and 4.

As was the case with the previous analyses, there was a general marked increase in desirable teacher behavior the third year of teaching. However, elementary teachers generally tended to stabilize or continue to improve from the third year to the fifth year. Secondary teachers appeared to stabilize or decline in desirable behavior between the third and fifth years. These observations are highly speculative and were not found statistically significant and are mentioned only as possible trends for continued study.

These findings from the COR analyses seem to establish a need for more support of the first year teacher in transitioning from preservice teacher education programs into the profession as first year teachers. Administrators and faculty in the college of education at Western Kentucky University are presently considering these data in deciding future directions for inservice and advanced degree programs in teacher education.

From Figures 2 and 3 there also appeared to be differences between elementary and secondary teachers for the various Phases. When t ratios were computed for COR variables in Phase 1, Phase 2, and Phase 4, significant differences were noted for Phase 1 and 2, but not for Phase 4. Phase 3 analyses were unavailable at the time of this writing due to technical problems with the computer. Table 8 presents these analyses. It appears that secondary student teachers and first year teachers are significantly below their elementary counterparts for all but COR Factor B. Phase 4 analyses did not yield significant differences for COR data between elementary and secondary participants.

Table 8

Differences Between Elementary and Secondary TPEP Participants on COR Factors

| COR | PHASE* 1 | | | 2 | | | 4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ | t | p | $\bar{x}$ | t | p | $\bar{x}$ | t | p |
| Factor A | | | | | | | | | |
| Elem. | 5.54 | 2.34 | .02 | 5.70 | 2.06 | .04 | 6.07 | .08 | .934 |
| Sec. | 5.27 | | | 5.36 | | | 6.10 | | |
| Factor B | | | | | | | | | |
| Elem. | 5.12 | .99 | .32 | 5.18 | .50 | .62 | 5.38 | .20 | .841 |
| Sec. | 4.99 | | | 5.08 | | | 5.41 | | |
| Factor C | | | | | | | | | |
| Elem. | 5.19 | 3.90 | .00 | 5.11 | 2.36 | .02 | 5.60 | 1.52 | .157 |
| Sec. | 4.66 | | | 4.65 | | | 4.77 | | |
| Factor D | | | | | | | | | |
| Elem. | 5.07 | 3.82 | .00 | 5.26 | 3.54 | .00 | 5.68 | .78 | .45 |
| Sec. | 4.53 | | | 4.51 | | | 5.31 | | |

* Phase 1 - Elementary n = 117, Secondary n = 106
Phase 2 - Elementary n = 90, Secondary n = 45
Phase 4 - Elementary n = 11, Secondary n = 7

Ten variables consisting of ratios from the CIAS were analyzed across Phases 2, 3, and 4. (See Appendix C for descriptions of CIAS variables.) The same repeated measures analyses of variance technique used for COR data was used for CIAS analyses. None of the teacher verbal behavior variables were found to differ significantly across Phases, indicating that teachers tend to remain consistent in their verbal behavior patterns from first through fifth years of teaching experience.

The ratio of "Student Questions to Student Responses" was significantly different across the three phases, but also approached significance in the interaction between Phase and Cycle. Most of these differences appear to be generated from the Cycle II Phase 2 mean of .324, a considerably larger value than the other means. Thus, the interpretation of this analysis is difficult and findings should be considered tentative until additional samples are studied. Table 9 contains this analysis.

"Student Talk to Total", the proportion of student talk, increased substantially from the first year of teaching to the third and fifth years of teaching. It would appear that more student verbal behavior was encouraged as the teachers became more experienced. Similar findings were found for the variable "Student Talk to Teacher Talk" and probably reflect the same increase in student verbal behavior. Table 10 and Table 11 contain the results of these analyses, respectively.

"Silence to Total", the proportion of silence categories recorded, decreased from Phase 2 to Phase 3 and Phase 4. The silence category most often reflects non-verbal directed activity. These findings are subject to bias from the presence of the observer in the classroom and may not be typical classroom behavior. However, the author feels strongly that in the large majority of the classrooms visited, the observer bias has been minimal.

Table 9

Repeated Measures for Elementary Variable
Student Questions to Student Responses

| CYCLE | n | PHASE 1 | 2 | 3 | 4 | $\bar{X}$ | Probability Phase | Cycle | Interaction |
|---|---|---|---|---|---|---|---|---|---|
| I | 6 | | .084 | .039 | .047 | .057 | .010 | .100 | .052 |
| II | 4 | | .324 | .048 | .034 | .135 | | | |
| $\bar{X}$ | 10 | | .180 | .043 | .042 | .088 | | | |

Table 10

Repeated Measures for Elementary Variable
Student Talk To Total

| CYCLE | n | PHASE 1 | 2 | 3 | 4 | $\bar{X}$ | Probability Phase | Cycle | Interaction |
|---|---|---|---|---|---|---|---|---|---|
| I | 6 | | .229 | .376 | .317 | .307 | .025 | .837 | .656 |
| II | 4 | | .245 | .341 | .361 | .316 | | | |
| $\bar{X}$ | 10 | | .235 | .362 | .335 | | | | |

Table 11

Repeated Measures for Elementary Variable
CIAS Student Talk to Teacher Talk

| CYCLE | n | PHASE 1 | 2 | 3 | 4 | $\bar{X}$ | Probability Phase | Cycle | Interaction |
|---|---|---|---|---|---|---|---|---|---|
| I | 6 | | .475 | .809 | .544 | .609 | .059 | .930 | .569 |
| II | 4 | | .403 | .740 | .716 | .620 | | | |
| $\bar{X}$ | 10 | | .446 | .781 | .613 | | | | |

Table 12

Repeated Measures for Elementary Variable
Silence to Total

| CYCLE | n | PHASE 1 | 2 | 3 | 4 | $\bar{X}$ | Probability Phase | Cycle | Interaction |
|---|---|---|---|---|---|---|---|---|---|
| I | 6 | | .276 | .125 | .096 | .166 | .035 | .456 | .244 |
| II | 4 | | .216 | .208 | .149 | .191 | | | |
| $\bar{X}$ | 10 | | .252 | .158 | .117 | .176 | | | |

Summary

The intent of this paper has been to introduce the Western Kentucky University Teacher Preparation Evaluation Program as one model of evaluating teacher education programs through the systematic follow-up of graduates. New findings from teacher effectiveness research and advancement in observational and other measurement techniques have provided a larger more lucrative pool from which to develop evaluation systems than was available in 1972. Undoubtedly, a somewhat different set of instruments and data gathering techniques would emerge if the model were revised in light of the new knowledge base. The important issue is not the specific instrumentation in use by TPEP; rather, the importance should be on a system which provides for objective, longitudinal data be collected from multiple sources within a research/conceptual framework. Such evaluation systems are essential if teacher education programs are to improve and provide products to meet the demands of contemporary education.

## References


Adams, R. D. "The Teacher Evaluation by a Peer/Supervisor II", Bowling Green, Kentucky: Unpublished Technical report from the Teacher Preparation Evaluation Program, 1978.

Adams, R. D. and Sandefur, J. T. "Factors Influencing the Perceived Problems of First Year Teachers," New Orleans, Louisiana: Paper presented at Mid South Educational Research Association, November, 1976.

Flanders, N. A. Teacher Influence Pupil Attitudes and Achievement (Cooperative Research Project No. 397) Washington, D.C.: U.S. Office of Health, Education and Welfare, Office of Education, 1960.

Haak, R. A., Kleiber, D. A., and Peck, R. F. Student Evaluation of Teacher II. Austin, Texas: Research and Development Center, University of Texas at Austin, 1970.

Nie, N. H., et. al. Statistical Package for the Social Sciences. New York, N.Y. McGraw-Hill, second edition, 1975.

Pegues, W. W. An Assessment of Teacher Education Follow-up Evaluation in the United States: A Descriptive Analysis. Tulsa, Oklahoma: Unpublished dissertation, The University of Tulsa, 1978.

Ryans, D. G. Characteristics of Teachers. Washington, D.C.: American Council on Education, 1960.

Sampson, P. B. "Analysis of Variance and Covariance including Repeated Measures" Bio Medical Computer Programs, Dixon, W. J. editor. Berkley, California: University of California Press, 1975.

Sandefur, J. T. An Illustrated Model for the Evaluation of Teacher Education Graduates. Washington, D.C.: American Association of Colleges for Teacher Education, 1972.

Shiek, D. A., and Adams, R. D. Factor Score Interpretation of the Classroom Observation Record. Bowling Green, KY: Western Kentucky University, Office of Educational Research, Technical Report, 1978.

Standards for Accreditation of Teacher Education. Washington, D.C.: National Council for Accreditation of Teacher Education: second edition, 1977.

Standards for State Approval of Teacher Education. Salt Lake City, Utah: National Association of State Directors of Teacher Education and Certification: sixth edition, 1976.

Veldman, Donald J. "Student Evaluation of Teaching": (Research Methodology Monograph No. 10) Austin, Texas: Research and Development Center, University of Texas, 1970.

# APPENDIX A

Table 1

Repeated Measures for Elementary Variable COR A Phase 1-3 Cycle 1-4

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 10 | $\bar{X}$ | 5.300 | 5.250 | 6.16 | 5.570 |
| | | SD | .793 | .871 | .453 | |
| II | 6 | $\bar{X}$ | 5.583 | 4.983 | 5.983 | 5.517 |
| | | SD | 1.096 | .688 | .574 | |
| III | 13 | $\bar{X}$ | 5.469 | 5.854 | 6.279 | 5.864 |
| | | SD | .554 | 1.064 | .330 | |
| IV | 10 | $\bar{X}$ | 5.520 | 5.760 | 5.960 | 5.747 |
| | | SD | .873 | .580 | .589 | |
| $\bar{X}$ | 39 | | 5.456 | 5.541 | 6.118 | 5.705 |

Main effects for Phase .000
Main effects for Cycle .335
Interaction effects .349

Table 2

Repeated Measures for Secondary Variable COR A Phase 1-3 Cycle 1-4

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 4.967 | 5.333 | 6.533 | 5.611 |
| | | SD | .929 | 1.301 | .058 | |
| II | 5 | $\bar{X}$ | 4.700 | 5.100 | 6.100 | 5.300 |
| | | SD | 1.158 | .791 | .292 | |
| III | 8 | $\bar{X}$ | 5.137 | 4.775 | 5.325 | 5.079 |
| | | SD | .560 | 1.585 | .667 | |
| IV | 3 | $\bar{X}$ | 5.200 | 5.633 | 5.167 | 5.333 |
| | | SD | .954 | .252 | 1.332 | |
| $\bar{X}$ | 19 | | 5.005 | 5.084 | 5.695 | 5.261 |

Main effects for Phase .083
Main effects for Cycle .523
Interaction effects .481

Table 3

Repeated Measures for Elementary Variable COR B Phase 1,2,3, Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 10 | $\bar{X}$ | 4.310 | 4.400 | 5.470 | 4.727 |
| | | SD | 1.137 | 1.275 | 1.361 | |
| II | 6 | $\bar{X}$ | 5.017 | 4.500 | 5.600 | 5.039 |
| | | SD | 1.179 | 1.658 | .890 | |
| III | 13 | $\bar{X}$ | 5.308 | 5.269 | 5.623 | 5.400 |
| | | SD | 1.010 | 1.527 | .878 | |
| IV | 10 | $\bar{X}$ | 5.120 | 4.920 | 5.280 | 5.107 |
| | | SD | 1.140 | 1.427 | 1.257 | |
| $\bar{X}$ | 39 | | 4.959 | 4.838 | 5.492 | 5.097 |

Main effects for Phase .004
Main effects for Cycle .419
Interaction effects .564

Table 4

Repeated Measures for Secondary Variable COR B Phase 1,2,3 Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.933 | 5.300 | 6.300 | 5.178 |
| | | SD | .929 | 1.300 | .436 | |
| II | 5 | $\bar{X}$ | 4.340 | 4.900 | 5.720 | 4.987 |
| | | SD | .691 | .778 | .507 | |
| III | 8 | $\bar{X}$ | 4.575 | 4.863 | 5.075 | 4.837 |
| | | SD | .822 | 1.511 | .573 | |
| IV | 3 | $\bar{X}$ | 5.700 | 5.133 | 5.133 | 5.322 |
| | | SD | .529 | 1.222 | 1.779 | |
| $\bar{X}$ | 19 | | 4.589 | 4.984 | 5.447 | 5.007 |

Main effects for Phase .027
Main effects for Cycle .689
Interaction effects .189

Table 5

Repeated Measures for Elementary Variable COR C Phases 1,2,3 Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 10 | $\bar{X}$ | 4.850 | 4.670 | 5.460 | 4.993 |
| | | SD | 1.288 | 1.137 | .659 | |
| II | 6 | $\bar{X}$ | 5.267 | 4.133 | 5.400 | 4.933 |
| | | SD | 1.067 | .863 | .844 | |
| III | 13 | $\bar{X}$ | 5.138 | 5.431 | 5.654 | 5.408 |
| | | SD | .853 | 1.447 | .711 | |
| IV | 10 | $\bar{X}$ | 5.300 | 4.730 | 5.100 | 5.043 |
| | | SD | .906 | 1.099 | .801 | |
| $\bar{X}$ | 39 | | 5.126 | 4.856 | 5.423 | 5.135 |

Main effects for Phase .006
Main effects for Cycle .436
Interaction effects .205

Table 6

Repeated Measures for Secondary Variable COR C Phase 1,2,3 Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.800 | 4.900 | 5.267 | 4.656 |
| | | SD | 1.732 | .755 | .252 | |
| II | 5 | $\bar{X}$ | 3.640 | 3.920 | 4.920 | 4.160 |
| | | SD | 1.050 | .507 | .698 | |
| III | 8 | $\bar{X}$ | 4.525 | 4.537 | 4.213 | 4.425 |
| | | SD | .891 | 1.514 | .476 | |
| IV | 3 | $\bar{X}$ | 5.500 | 5.067 | 4.167 | 4.911 |
| | | SD | .529 | .462 | 1.930 | |
| $\bar{X}$ | 19 | | 4.332 | 4.516 | 4.558 | 4.468 |

Main effects for Phase .698
Main effects for Cycle .388
Interaction effects .131

Table 7

Repeated Measures for Elementary Variable COR D Phases 1,2,3 Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 10 | $\bar{X}$ | 4.740 | 4.630 | 4.920 | 4.763 |
| | | SD | 1.064 | 1.148 | 1.377 | |
| II | 6 | $\bar{X}$ | 4.817 | 4.167 | 5.533 | 4.839 |
| | | SD | 1.055 | 1.157 | .933 | |
| III | 13 | $\bar{X}$ | 5.415 | 5.362 | 5.523 | 5.433 |
| | | SD | .848 | 1.289 | .778 | |
| IV | 10 | $\bar{X}$ | 4.720 | 5.240 | 5.280 | 5.080 |
| | | SD | 1.202 | 1.302 | 1.093 | |
| $\bar{X}$ | 39 | | 4.972 | 4.959 | 5.308 | 5.079 |

Main effects for Phase .091
Main effects for Cycle .219
Interaction effects .471

Table 8

Repeated Measures for Secondary Variable COR D Phase 1,2,3 Cycle I,II,III,IV

| | | | PHASE | | | |
|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.767 | 5.200 | 6.200 | 5.056 |
| | | SD | 1.102 | 1.353 | .361 | |
| II | 5 | $\bar{X}$ | 3.600 | 4.540 | 5.280 | 4.473 |
| | | SD | .834 | .896 | .769 | |
| III | 8 | $\bar{X}$ | 4.437 | 3.375 | 4.725 | 4.179 |
| | | SD | 1.472 | 1.233 | .696 | |
| IV | 3 | $\bar{X}$ | 4.053 | 4.200 | 4.767 | 4.556 |
| | | SD | .814 | 1.234 | 1.563 | |
| $\bar{X}$ | 19 | | 4.053 | 4.200 | 5.111 | 4.454 |

Main effects for Phase .006
Main effects for Cycle .320
Interaction effects .106

# APPENDIX B

Table 1

Repeated Measures for Elementary Variable COR A Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 6 | $\bar{X}$ | 5.517 | 5.450 | 6.333 | 6.000 | 5.825 |
| | | SD | .788 | .568 | .403 | .473 | |
| II | 4 | $\bar{X}$ | 5.475 | 4.925 | 5.775 | 5.975 | 5.537 |
| | | SD | 1.360 | .866 | .591 | .613 | |
| $\bar{X}$ | 10 | | 5.500 | 5.240 | 6.110 | 5.990 | 5.710 |

Main effects for Phase .058
Main effects for Cycle .180
Interaction effects .780

Table 2

Repeated Measures for Secondary Variable COR A Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 4.967 | 5.333 | 6.533 | 6.367 | 5.800 |
| | | SD | .929 | 1.301 | .058 | .681 | |
| II | 4 | $\bar{X}$ | 4.950 | 4.925 | 6.150 | 5.900 | 5.481 |
| | | SD | 1.170 | .793 | .311 | .804 | |
| $\bar{X}$ | 7 | | 4.957 | 5.100 | 6.314 | 6.100 | 5.618 |

Main effects for Phase .017
Main effects for Cycle .369
Interaction effects .959

Table 3

Repeated Measures for Elementary Variable COR B Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 6 | $\bar{X}$ | 4.200 | 4.450 | 5.400 | 5.050 | 4.775 |
| | | SD | 1.365 | 1.190 | 1.663 | 1.078 | |
| II | 4 | $\bar{X}$ | 4.825 | 3.925 | 5.375 | 5.875 | 5.000 |
| | | SD | 1.226 | .513 | .862 | .822 | |
| $\bar{X}$ | 10 | | 4.450 | 4.240 | 5.390 | 5.380 | 4.865 |

Main effects for Phase .020
Main effects for Cycle .685
Interaction effects .440

Table 4

Repeated Measures for Secondary Variable COR B Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.933 | 5.300 | 6.300 | 5.533 | 5.267 |
| | | SD | .929 | 1.300 | .436 | 1.305 | |
| II | 4 | $\bar{X}$ | 4.525 | 4.675 | 5.750 | 5.325 | 5.069 |
| | | SD | .640 | .685 | .580 | .998 | |
| $\bar{X}$ | 7 | | 4.271 | 4.943 | 5.986 | 5.414 | 5.154 |

Main effects for Phase .030
Main effects for Cycle .175
Interaction effects .667

Table 5

Repeated Measures for Elementary Variable COR C Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 6 | $\bar{X}$ | 5.167 | 4.967 | 5.483 | 5.250 | 5.217 |
| | | SD | .423 | .9136 | .794 | 1.126 | |
| II | 4 | $\bar{X}$ | 4.925 | 4.125 | 5.275 | 5.975 | 5.075 |
| | | SD | .974 | .838 | .918 | .714 | |
| $\bar{X}$ | 10 | | 5.070 | 4.630 | 5.400 | 5.540 | 5.160 |

Main effects for Phase .060
Main effects for Cycle .630
Interaction effects .272

Table 6

Repeated Measures for Secondary Variable COR C Phases 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.800 | 4.900 | 5.267 | 5.567 | 4.883 |
| | | SD | 1.732 | .755 | .252 | .723 | |
| II | 4 | $\bar{X}$ | 3.825 | 3.775 | 4.775 | 4.175 | 4.137 |
| | | SD | 1.115 | .450 | .714 | 1.417 | |
| $\bar{X}$ | 7 | | 3.814 | 4.257 | 4.986 | 4.771 | 4.457 |

Main effects for Phase .099
Main effects for Cycle .178
Interaction effects .493

Table 7

Repeated Measures for Elementary Variable COR D Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 6 | $\bar{X}$ | 5.150 | 4.900 | 5.033 | 5.333 | 5.104 |
| | | SD | .907 | .678 | 1.729 | .314 | |
| II | 4 | $\bar{X}$ | 4.575 | 3.800 | 5.275 | 6.075 | 4.931 |
| | | SD | 1.274 | .942 | 1.069 | 1.090 | |
| $\bar{X}$ | 10 | | 4.920 | 4.460 | 5.130 | 5.630 | 5.035 |

Main effects for Phase .026
Main effects for Cycle .718
Interaction effects .149

Table 8

Repeated Measures for Secondary Variable COR D Phase 1-4 Cycle 1-2

| | | | PHASE | | | | |
|---|---|---|---|---|---|---|---|
| CYCLE | n | | 1 | 2 | 3 | 4 | $\bar{X}$ |
| I | 3 | $\bar{X}$ | 3.767 | 5.200 | 6.200 | 5.833 | 5.250 |
| | | SD | 1.102 | 1.353 | .361 | .569 | |
| II | 4 | $\bar{X}$ | 3.500 | 4.300 | 5.150 | 4.925 | 4.469 |
| | | SD | .927 | .829 | .823 | 1.297 | |
| $\bar{X}$ | 7 | | 3.614 | 4.686 | 5.600 | 5.314 | 4.804 |

Main effects for Phase .008
Main effects for Cycle .074
Interaction effects .884

# APPENDIX C

# CIAS Category Description

| | Category | Description of Behavior |
|---|---|---|
| | 1 | Accepts Feeling: Accepts and clarifies the feeling and tone of students in a non-threatening manner. Feelings may be positive or negative. Predicting and recalling feelings are also included. |
| | 2 | Praises or Encourages: Praises or encourages student action or behavior. Jokes that release tension not at the expense of another individual, saying "uh-huh" or "go on" are included. |
| T | 3 | Accepts or Uses Students' Ideas: Clarifying, building on, developing, and accepting ideas of students. |
| E | | |
| A | 4 | Asks Direct Questions: Asking a question about content or procedure with the intent the student should give a direct recall response. |
| C | 41 | Asks Probing Questions: Asking a question about content or procedure with the intent the student should give more information or explain further a statement or answer. |
| H | | |
| E | 42 | Asks Higher Order Questions: Asking a question about content or procedure with the intent the student should give an answer that requires opinion, analysis, synthesis, evaluation, etc. on the part of the student. |
| R | | |
| | 5 | Answers Student Questions: Answers to student questions regarding content or procedure. |
| T | | |
| A | 6 | Lecture Content: Giving facts or opinions about the subject or content of the class; expressing own ideas, asking rhetorical questions. |
| L | | |
| K | 61 | Lecture Non-Content: Giving information or making comments on procedures or topics other than content of class. |
| | 7 | Corrective Feedback: Telling a student directly that his answer or statement is incorrect. |
| | 8 | Gives Directions: Directions, commands, or orders to which a student is expected to comply immediately. Does not apply to information about procedures or future assignments. |
| | 9 | Criticizes or Justifies Authority: Direct statements of a critical nature intended to change student behavior from a non-acceptable to an acceptable pattern; bawling out the student; stating why the teacher is doing what he/she is doing so as to achieve or maintain control; rejecting or criticizing a student's opinion or judgment; using sarcasm. |
| | 10 | Student Talk Initiated By Teacher: Talk by students in response to requests or questions by teacher. The teacher initiates the contact or solicits student's statement. |
| S T U D E N T T A L K | 101 | Student Initiated Talk: Talk initiated by students, not in response to teacher talk. |
| | 11 | Student Questions: Questions concerning content or procedure that are directed to the teacher or other students. |
| S | 12 | Directed Practice or Activity: Non-verbal behavior requested or suggested by the teacher. Seat work and small groups work are in this category. |
| I | | |
| L | 13 | Demonstration: Silence during periods when visual materials are being shown or when non-verbal demonstration is being conducted by the teacher or by a student. |
| E | | |
| N | 14 | Constructive Silence: Silence during the classroom that is constructive such as "think time" or time given a student to respond to a question. |
| C | | |
| E | 141 | Non-Constructive Silence and Confusion: Periods of confusion where observer cannot understand communication, other special classroom activities not covered in above category. |

## CIAS Variable Description

1. Little i/d - Teacher behaviors coded in categories 1, 2, and 3 were divided by teacher behaviors coded in categories 7, 8, and 9 to form a "pure" measure of indirect (pupil oriented/supportive focus) to direct (teacher oriented focus) teacher behavior.

2. Big I/D - Teacher behaviors coded in categories 1 through 5 were divided by teacher behaviors coded in categories 6 through 9 to form a more "gross" measure of indirect to direct teacher behavior that took into account questions and lecture.

3. Student Talk/Teacher Talk (ST TLK/TCH TLK) - The ratio between the amount of student verbal behavior and teacher verbal behavior.

4. Silence/Total (Sil/Tot) - The proportion of silence recorded in the classroom where silence was constituted of categories 12, 13, 14, 141.

5. Lecture/Total (Lec/Tot) - The proportion of lecture recorded in the classroom; category 6 and 61 constituted lecture.

6. Indirect/Total (Ind/Tot) - The proportion of "pure" indirect teacher behavior recorded (categories 1, 2, 3)

7. Direct/Total (Dir/Tot) - The proportion of "pure" direct teacher behavior recorded (categories 7, 8, 9)

8. Teacher Talk/Total - The proportion of teacher verbal behavior recorded.

9. Student Questions/Student Responses (StQ/StR) - Ratio of student questions (category 11) to student responses (categories 10, 101).

10. Student Talk/Total - Proportion of student verbal behavior.

The cost of printing this publication by Western Kentucky University was paid from state funds KRS 57.375.

# COMMENTS FROM THE SYMPOSIUM DISCUSSANTS

Karen B. Kepler

Teachers College
Columbia University

Robert F. Peck

University of Texas at Austin

KAREN B. KEPLER, DISCUSSANT REMARKS:

All the presenters should be congratulated for giving more than lip service to accreditation demands for evaluation. I think that they are right that their ambitious efforts are not usual in our field.

As a critic, I toyed with the idea of following the model of evaluation that they used in their papers, asking all of you, all of them, and myself to rate them on a 1-9 scale in terms of the symposium objectives. I decided not to do that.

The symposium objectives were: to offer tools and techniques, report selected findings, and raise issues to be considered by those initiating studies. They do offer a variety of tools and techniques as well as uses of evaluation. The selected findings are appropriately slim -- since most of the results of such efforts are so idiosyncratic as to be of interest primarily to the institution concerned. As a group, my major disappointment is that the "issues to be considered by those initiating studies" did not receive the priority I feel it deserves. I would like to highlight some of these issues as I discuss each paper.

There seemed to be at least two distinct types of evaluation functions demonstrated by these studies. The Schalock, Adams and part of the Ayers papers focus on evaluation for prediction purposes. The Cooper/Jones and Ayers papers use evaluation for program improvement purposes. And, the Ryan paper, not an evaluation study, draws implications for evaluation studies from a descriptive study of first year teachers. Let me first look at the prediction studies.

The Adams and Ayers papers present all too rare systematically collected longitudinal data. They essentially follow the same design, collecting data using a variety of methods during student teaching, the first, third and fifth year of teaching. In these particular papers, Ayers talks about predicting success and Adams about predicting attrition from teaching. For his sample, Ayers found that the less authoritarian people were better teachers and that the NTE and Grade Point Average in ed and psych courses were the best predictors of later success. In reading the paper, I asked myself what does best mean? How much do they actually predict? In contrast, Adams finds that attrition is not related to any of the many variables that he studied, including success. (As an aside, in his paper he says it is distressing that 8 percent of the people did not want to go into teaching. I don't think that is a disappointingly large number at all. I think it is rather small. I don't think our prediction and vocational guidance are good enough at this point that 92 percent of the people know what they are doing when they enter teacher education programs. So, I am not disturbed by the 8 percent figure at all.) Anyway, Adams concludes that the findings suggest a more complex set of factors operating than he studied. I agree. Given Ryan's findings about the situational and personal factors affecting first year teachers, one wonders whether the phenomenon will ever yield to neat prediction studies. Perhaps, their value is to continually remind us of the complexity of our task and keep us from labelling students prematurely.

I think that this is what troubles me most about the clear-cut rational model

presented by the Schalock paper. It is too simplistically appealing. His aim is to find out who is likely to have the most success in achieving desired, learning outcomes with particular children under particular learning conditions. It is no easy task to, as he says, "statistically or experimentally control related sources of variance" or "measure factors influencing measures of effectiveness." The research problem here seems to me more difficult than the selection problem it is supposed to help aid. Besides being difficult, I doubt whether such fine-tuned prediction is worth the time and energy given the realities of teacher selection, placement, attrition and student grouping. Even if possible, such an approach seems a little too deterministic, minimizing individual diversity, judgment and growth.

Another problem -- which is underplayed in all three papers -- is a determination of measures of effectiveness. Perhaps, because of time and space limitations, there is an emphasis on technique and instrumentation, rather than on the difficulties and criteria used in determining what one is going to measure. For initiators of follow-up studies, I think this issue is far more problemmatic than tools and techniques. If it isn't, then we do have a problem.

The part of the Adams and Ayers studies which I think are most useful are the descriptions of change over time. They come up with different results, perhaps reflecting differences in what they are measuring, as much as sample and program differences. Data from such longitudinal studies can uniquely contribute to our understanding of teaching, for they say as much about the nature of the job as the adequacy of the person or the program. And, it is particularly useful for teacher education programs which are now switching to the emphasis from preservice to inservice. I suspect there is a limit to the amount that we can accomplish in a preservice program any how -- trying to anticipate and remedy all the turmoils facing the maturing teachers is an unrealistic demand to place on ourselves and on our students. Certainly, however, these are fertile fields for program development and inservice education.

Now the studies which were specifically designed for program improvement: the Cooper/Jones paper and the studies Ayers described. Both institutions apparently have done an unusually good job of making evaluation an integral part of program development. In looking at these studies, I would like to focus on two related issues -- one being the evaluation of objectives, and the other being the use of data once it is collected.

Cooper/Jones state the truism which many evaluators, rushing about to find instrumentation often forget -- "the developers of each program must decide what questions they want to answer before collecting any information." I think their study is a model of clarity here, although I would have been more interested in discrepancy evaluation which also included content as well as structural characteristics. My main concern, however, is that in the evaluation process the objectives may have become written in stone. For instance, they report great concern that the foundations courses were not modularized or not competency-based, which is one of the objectives that they were evaluating. I think part of the role of evaluators is not just to measure the attainment of given objectives, but also to assess their appropriateness. I think an argument could be made that a competency-based modularized approach is not the most appropriate strategy for all goals -- foundations understandings perhaps being better achieved and

measured in other ways.

Interestingly, I thought the Ayers study was going to avoid this trap when one of his stated objectives was "evaluating the objectives of the teacher education program." Reading on, one finds that he means the attainment of the objectives rather than the appropriateness of the objectives themselves. I think educational evaluators should be more than technicians evaluating given objectives, but also signal and participate in the re-development of the objectives themselves.

In the future I would also like to see more attention to how this data is used once it is collected. Ayers does give us a list of 14 different evaluation studies and 27 resulting programmatic changes. That is quite impressive. But while he details procedures, what is missing for the new initiator is a description of the process of deliberation. For instance, while it may be obvious what to do to remedy the deficiency of "cognitive knowledge of science," what does a faculty decide when told -- once more -- that classroom control is still a deficiency of its graduates. Is this amenable to instruction? What kind that has not already been attempted? Or is this a developmental issue more subject to maturation and experience? Where should our energies be directed as teacher educators? The complexity of the utilization of data is even clearer in the Cooper/Jones results because they looked at more complex things, I think. First of all, I am surprised that they are surprised that the university supervisor rated the student performance lowest and that the student teacher rated their own performance the highest. Not only does the university supervisor have more relative data but is also more likely to hold to ideals which may or may not be attainable in day to day classroom teaching. Other differential perceptions can also be explained in the reality base of practitioners, the lack of knowledge of the university supervisor who comes and observes a few selected lessons and the particular concerns of one being evaluated. Now what does one do? Accept these different perceptions? Or try to achieve greater congruency? And, if so, in what direction? And, what happens to a competency which is voted low in importance and performance? For example, incorporating the cultural experiences of children into curriculum planning was rated low by everybody in terms of importance and also low in terms of student performance. What do you do? Try to change the perceptions of people in terms of their priorities, work on the performance or put it low down and work on another objective which everybody seemed to rate high in importance but low in performance, such as incorporating principles of human growth and development into the curriculum. I think collecting data is only the first part of follow-up studies for curriculum improvement. We need as much attention to articulation of results into program as to collection techniques if these studies are to serve their purposes.

The last study, the Ryan study, is the least rigorous study and is not an evaluation study at all, but I think has some of the strongest implications for follow-up studies. The instability and unpredictability of first year teachers' performance and emotion, the highly subjective self evaluation, the non-contingent renewal of contracts, and the impact of situational and personal contexts highlight the difficulties inherent in program effectiveness studies. Ryan also gives us an alternative source of content for evaluation studies -- the realities of first year teaching -- in contrast with the competencies to be

achieved in a preservice program. Perhaps we should be evaluating different competencies and dilemmas at different levels of teaching. One caveat here -- basing objectives solely upon analysis of job/life demands could lead to the narrowness of a Bobbitt curriculum and an unexamined perpetuation of the status quo.

Ryan's study also points to another kind of finding of central importance -- these first year teachers valued the non-evaluative attention of these descriptive researchers. This finding does not reflect the deficiency of any program, but does rather pinpoint a critical problem in the structure of teacher education in this country, and that is that we drop them after the preservice year and that they usually do not have any attention from anybody their first year. This is a clear example that follow-up studies need to be broadly conceived -- not limited to prediction studies nor studies assessing attainment of competency objectives -- if we are to move forward rather than being trapped in present modes of operation.

ROBERT F. PECK, DISCUSSANT REMARKS:

I would just like to make a few bridging observations which I think I detect in the things these folks have been doing for some years. First, the examples of the Houston program and the Tennessee program, in particular, are almost unique, at least to my knowledge, in using follow-up data to try to modify their pre-service programs. Everybody talks about it but nobody does it. They have done it. They can serve any institution, I think, as a good model, of how to start getting your faculty using some kind of data to improve their instructional programs. The divorce between evaluation, and program design and execution can hardly be overstated; in most places, they never meet. In these two places they have met. These people have figured out some ways of getting some of the members of their faculty to get less defensive and participate more actively. They said some, not all, I know, but that would be true anywhere. They have carefully designed sets of data which can be used for evaluation and feedback. Jim's paper described a good feedback process in some detail. Rather than talk about the program broadly, he says, look at specific components within that program, asking very specific questions about what we say we are doing. Are we doing it, first of all? Next, does anybody care? Then you go back to the drawing board and either reassess the priority you have given that component or reassess the degree of excellence with which you are approaching it, or a little of both.

One of the difficulties I found with Del's paper (this is probably unfair to him, since he condensed his thinking into a very brief span for his presentation) is that I thought I heard the implication that being hired is a good criterion of effective teaching. In at least two of these studies, both in Ron's findings and in Kevin's very careful observations, that is not true. Why people get jobs teaching and why they keep them does not have a great deal to do with the effectiveness of their teaching behavior. "Effective" by what criterion? They use various approaches, but both the Tennessee and the Kentucky programs are using Ryan's Classroom Observation Record. I'm biased here. I have been using it, too, for the same reasons they are: about ten or twelve years ago J. T. Sandefur was able to demonstrate that the Ryan's scales related very significantly to other independent measures of teaching effectiveness, and were affected by training programs. These scales seem to me to be about the right level of generality to transcend the minute differences in the teaching situations and in individual pupil differences, so that you find a good deal of stability of individual teacher behavior across time and across samples, using that kind of measure. Having used this system, I would concur with these people that this is about as close as you can come, in an economical fashion, to capture the essence of what a teacher is doing that is likely to make an important difference with children.

This has to be said in just this cautious, probabilistic way. We have very, very difficult problems creating good designs for measuring teacher effects on pupil changes. One of the problems already alluded to here is that situational factors make a tremendous difference in how any given teacher affects any given child or group of children. It's a little bit buried in Ron's brief recitation and in Jerry's, but it is quite explicit in Kevin's. Kevin said, the individual teacher's situation in each case is unique, and it powerfully affects the teacher's

behavior, morale and attitudes. If you stop and think a moment, you will realize that any of us who are in the preservice business are training people who will fan out to the most unbelievably diverse array of places, communities, school systems, with different kinds of principals, different kinds of teacher groups, all of which have a very strong influence on what a teacher does and how it affects the children. None of these contextual factors have been assessed in any of these studies, for good reasons. You can't do that without a lot more money than anybody has come up with yet. All of these studies are subject to the caveat that the teachers are differently located. Five years after graduation, 31 teachers are likely to be in 30 different communities, at eight or ten different grade levels.

Kevin's observation also is vitally important, that the morale, the energy level, and the teaching behavior of each teacher seems to be strongly affected by what is happening in his or her private life. He gives vivid examples: a death in the family, a divorce, a robbery and rape attempt. Such experiences powerfully change the way a teacher comes into a classroom, and what he or she does during the following days and weeks.

The bottom line of all this, I think, is the one that Cronbach made five years ago. We need to take a case study approach. The facts require it. We need to take a very intensive, studious collection of information about all of these relevant data about each teacher. We cannot afford to measure large samples of teachers in this way with present funds; but we can, in a given place, come to a very accurate and insightful picture of what is going on in that place. After that, we can worry about creating typologies, classifying subsets of patterns, and seeing how they go together to make somewhat more general "laws" of teacher behavior.